

THE  
INTERNATIONAL SERIES  
OF  
MONOGRAPHS ON PHYSICS

GENERAL EDITORS

R. H. FOWLER AND P. KAPITZA

# THE INTERNATIONAL SERIES OF MONOGRAPHS ON PHYSICS

## GENERAL EDITORS

R. H. FOWLER

F.R.S., Fellow of Trinity College, Cambridge, Professor of Applied Mathematics in the University of Cambridge

P. KAPITZA

F.R.S., Fellow of Trinity College, Cambridge, Royal Society Messel Research Professor

## *Already Published*

CONSTITUTION OF ATOMIC NUCLEI AND RADIOACTIVITY. By G. GAMOW. 1931. Royal 8vo, pp. 122.

THE THEORY OF ELECTRIC AND MAGNETIC SUSCEPTIBILITIES. By J. H. VAN VLECK. 1932. Royal 8vo, pp. 396.

WAVE MECHANICS. ELEMENTARY THEORY. By J. FRENKEL. 1932. Royal 8vo, pp. 286.

WAVE MECHANICS. ADVANCED GENERAL THEORY. By J. FRENKEL. 1934. Royal 8vo, pp. 533.

THE THEORY OF ATOMIC COLLISIONS. By N. F. MOTT and H. S. W. MASSEY. 1933. Royal 8vo, pp. 300.

RELATIVITY, THERMODYNAMICS, AND COSMOLOGY. By R. C. TOLMAN. 1934. Royal 8vo, pp. 518.

ELECTROLYTES. By HANS FALKENHAGEN. Translated by R. P. BELL. 1934. Royal 8vo, pp. 364.

CHEMICAL KINETICS AND CHAIN REACTIONS. By N. SEMENOFF. 1935. Royal 8vo, pp. 492.



THE  
P R I N C I P L E S  
O F  
QUANTUM MECHANICS

BY  
P. A. M. D I R A C

FELLOW OF ST. JOHN'S COLLEGE  
CAMBRIDGE

*SECOND EDITION*

OXFORD  
AT THE CLARENDON PRESS  
1935

OXFORD UNIVERSITY PRESS  
AMEN HOUSE, E.C. 4  
LONDON EDINBURGH GLASGOW  
NEW YORK TORONTO MELBOURNE  
CAPETOWN BOMBAY CALCUTTA  
MADRAS SHANGHAI  
HUMPHREY MILFORD  
PUBLISHER TO THE UNIVERSITY

PRINTED IN GREAT BRITAIN



## PREFACE TO SECOND EDITION

THE book has been mostly rewritten. I have tried by carefully overhauling the method of presentation to give the development of the theory in a rather less abstract form, without making any sacrifices in exactness of expression or in the logical character of the development. This should make the work suitable for a wider circle of readers, although the reader who likes abstractness for its own sake may possibly prefer the style of the first edition.

The main change has been brought about by the use of the word 'state' in a three-dimensional non-relativistic sense. It would seem at first sight a pity to build up the theory largely on the basis of non-relativistic concepts. The use of the non-relativistic meaning of 'state', however, contributes so essentially to the possibilities of clear exposition as to lead one to suspect that the fundamental ideas of the present quantum mechanics are in need of serious alteration at just this point, and that an improved theory would agree more closely with the development here given than with a development which aims at preserving the relativistic meaning of 'state' throughout.

Some mistakes which have been kindly pointed out to me by friends have been corrected and some new subject-matter has been inserted, the largest addition being a chapter on field theory.

P. A. M. D.

THE INSTITUTE FOR ADVANCED STUDY,

PRINCETON.

27 November 1934.



## FROM THE PREFACE TO THE FIRST EDITION

THE methods of progress in theoretical physics have undergone a vast change during the present century. The classical tradition has been to consider the world to be an association of observable objects (particles, fluids, fields, etc.) moving about according to definite laws of force, so that one could form a mental picture in space and time of the whole scheme. This led to a physics whose aim was to make assumptions about the mechanism and forces connecting these observable objects, to account for their behaviour in the simplest possible way. It has become increasingly evident in recent times, however, that nature works on a different plan. Her fundamental laws do not govern the world as it appears in our mental picture in any very direct way, but instead they control a substratum of which we cannot form a mental picture without introducing irrelevancies. The formulation of these laws requires the use of the mathematics of transformations. The important things in the world appear as the invariants (or more generally the nearly invariants, or quantities with simple transformation properties) of these transformations. The things we are immediately aware of are the relations of these nearly invariants to a certain frame of reference, usually one chosen so as to introduce special simplifying features which are unimportant from the point of view of general theory.

The growth of the use of transformation theory, as applied first to relativity and later to the quantum theory, is the essence of the new method in theoretical physics. Further progress lies in the direction of making our equations invariant under wider and still wider transformations. This state of affairs is very satisfactory from a philosophical point of view, as implying an increasing recognition of the part played by the observer in himself introducing the regularities that appear in his observations, and a lack of arbitrariness in the ways of nature, but it makes things less easy for the learner of physics. The new theories, if one looks apart from their mathematical setting, are built up from physical concepts which cannot be explained in terms of things previously known to the student, which cannot even be explained adequately in words at all. Like the fundamental concepts (e.g. proximity, identity) which every one must learn on his



arrival into the world, the newer concepts of physics can be mastered only by long familiarity with their properties and uses.

From the mathematical side the approach to the new theories presents no difficulties, as the mathematics required (at any rate that which is required for the development of physics up to the present) is not essentially different from what has been current for a considerable time. Mathematics is the tool specially suited for dealing with abstract concepts of any kind and there is no limit to its power in this field. For this reason a book on the new physics, if not purely descriptive of experimental work, must be essentially mathematical. All the same the mathematics is only a tool and one should learn to hold the physical ideas in one's mind without reference to the mathematical form. In this book I have tried to keep the physics to the forefront, by beginning with an entirely physical chapter and in the later work examining the physical meaning underlying the formalism wherever possible. The amount of theoretical ground one has to cover before being able to solve problems of real practical value is rather large, but this circumstance is an inevitable consequence of the fundamental part played by transformation theory and is likely to become more pronounced in the theoretical physics of the future.

With regard to the mathematical form in which the theory can be presented, an author must decide at the outset between two methods. There is the symbolic method, which deals directly in an abstract way with the quantities of fundamental importance (the invariants, etc., of the transformations) and there is the method of coordinates or representations, which deals with sets of numbers corresponding to these quantities. The second of these has usually been used for the presentation of quantum mechanics (in fact it has been used practically exclusively with the exception of Weyl's book *Gruppentheorie und Quantenmechanik*). It is known under one or other of the two names 'Wave Mechanics' and 'Matrix Mechanics' according to which physical things receive emphasis in the treatment, the states of a system or its dynamical variables. It has the advantage that the kind of mathematics required is more familiar to the average student, and also it is the historical method.

The symbolic method, however, seems to go more deeply into the nature of things. It enables one to express the physical laws in a neat and concise way, and will probably be increasingly used in the future as it becomes better understood and its own special mathematics gets



developed. For this reason I have chosen the symbolic method, introducing the representatives later merely as an aid to practical calculation. This has necessitated a complete break from the historical line of development, but this break is an advantage through enabling the approach to the new ideas to be made as direct as possible. I have given the connexion between the new theory and Bohr's orbit theory, because the latter is likely to be useful in an elementary way for a long time to come.

P. A. M. D.

ST. JOHN'S COLLEGE, CAMBRIDGE.

29 *May* 1930.



# CONTENTS

I. THE PRINCIPLE OF SUPERPOSITION	1
1. The Need for a Quantum Theory	1
2. The Polarization of Photons	4
3. Interference of Photons	7
4. Superposition and Indeterminacy	10
5. Mathematical Formulation of the Principle	14
6. Analysis of the Principle	17
II. STATES AND OBSERVABLES	20
7. The Vector Space representing the States	20
8. Observables as Linear Operators	24
9. Eigenvalues	30
10. The Expansion Theorem	34
11. Functions of an Observable	38
12. The General Physical Interpretation	42
13. Commutability and Compatibility	46
III. REPRESENTATION THEORY FOR DISCRETE EIGEN-VALUES	49
14. The Bracket Notation	49
15. Matrix Multiplication	52
16. Eigen- $\psi$ 's as Basic $\psi$ 's	54
17. Transformation Theory	60
18. Probability Amplitudes	65
19. Example	67
IV. REPRESENTATION THEORY FOR CONTINUOUS EIGEN-VALUES	71
20. Introduction of the $\delta$ function	71
21. Properties of the $\delta$ function.	74
22. Representations with One Continuous Parameter	77
23. General Representations	82
24. The Weight Function	85
V. THE QUANTUM CONDITIONS	88
25. Poisson Brackets	88
26. Canonical Coordinates and Momenta	92
27. Momenta as Differential Operators	95
28. Heisenberg's Principle of Uncertainty	101
29. Displacement Operators	104
30. Contact Transformations	109
VI. THE EQUATIONS OF MOTION	115
31. Schrödinger's Form for the Equations of Motion	115
32. Heisenberg's Form for the Equations of Motion	118
33. The Action Principle	123



34. The Motion of Wave Packets . . . . .	126
35. The Free Particle . . . . .	130
36. The Harmonic Oscillator . . . . .	133
37. The Gibbs Ensemble . . . . .	139
 VII. MOTION IN A CENTRAL FIELD OF FORCE . . . . .	145
38. Introduction of the Angular Momentum . . . . .	145
39. Properties of Angular Momentum . . . . .	147
40. Transition to Polar Coordinates . . . . .	151
41. Energy-levels of the Hydrogen Atom . . . . .	155
42. Selection Rules . . . . .	158
43. The Zeeman Effect for the Hydrogen Atom . . . . .	162
44. Combination of Angular Momenta . . . . .	164
 VIII. PERTURBATION THEORY . . . . .	168
45. General Remarks . . . . .	168
46. The Change in the Energy-levels caused by a Perturbation . . . . .	169
47. The Perturbation considered as causing Transitions . . . . .	173
48. Application to Radiation . . . . .	178
49. Transitions caused by Perturbation Independent of the Time . . . . .	180
50. The Anomalous Zeeman Effect . . . . .	183
 IX. COLLISION PROBLEMS . . . . .	187
51. General Remarks . . . . .	187
52. The Scattering Coefficient . . . . .	190
53. Solution with the $p$ -Representation . . . . .	195
54. Dispersive Scattering . . . . .	201
55. Resonance Scattering . . . . .	203
56. Emission and Absorption . . . . .	206
 X. SYSTEMS CONTAINING SEVERAL SIMILAR PARTICLES . . . . .	209
57. Symmetrical and Antisymmetrical States . . . . .	209
58. Permutations as Dynamical Variables . . . . .	213
59. Permutations as Constants of the Motion . . . . .	215
60. Determination of the Energy-levels . . . . .	219
61. Application to Electrons . . . . .	223
 XI. THEORY OF RADIATION . . . . .	229
62. Second Quantization . . . . .	229
63. Waves and Einstein-Bose Particles . . . . .	234
64. Application to Photons . . . . .	237
65. Determination of the Interaction Energy between a Photon and an Atom . . . . .	241
66. Emission, Absorption, and Scattering of Radiation . . . . .	244
67. Einstein's Laws of Radiation . . . . .	248
 XII. THE RELATIVISTIC THEORY OF THE ELECTRON . . . . .	251
68. Relativistic Treatment of a Particle . . . . .	251
69. The Wave Equation for the Electron . . . . .	252
70. Invariance under a Lorentz Transformation . . . . .	255



# CONTENTS

xi

71. The Motion of a Free Electron . . . . .	260
72. Existence of the Spin . . . . .	262
73. Transition to Polar Variables . . . . .	265
74. The Fine-Structure of the Energy-levels of Hydrogen . . . . .	267
75. Theory of the Positron . . . . .	270
<b>XIII. FIELD THEORY . . . . .</b>	<b>273</b>
76. Quantum Conditions for the Electromagnetic Field . . . . .	273
77. Quantum Conditions for the Electromagnetic Potentials . . . . .	278
78. The Supplementary Conditions . . . . .	282
79. Interaction of Field and Particles . . . . .	286
80. The Quantization of Electron Waves . . . . .	291
81. Conclusion . . . . .	296
<b>INDEX OF DEFINITIONS . . . . .</b>	<b>299</b>





## THE PRINCIPLE OF SUPERPOSITION

## 1. The Need for a Quantum Theory

CLASSICAL mechanics has been developed continuously from the time of Newton and applied to an ever-widening range of dynamical systems, including, after the formalism is adapted to relativity requirements, the electromagnetic field in interaction with matter. The underlying ideas and the laws governing their application form a simple and elegant scheme, which one might be inclined to think could not be seriously modified without having all its attractive features spoilt. Nevertheless the passage to a new scheme, called quantum mechanics, has been found to be necessary for the discussion of phenomena on the atomic scale and the new scheme is even more elegant and satisfying than the classical one. This is brought about by the fact that the changes which the new scheme requires are of a very profound character and do not clash with those shallower features that make the classical theory so attractive, as a result of which all these features can be taken over unchanged into the new scheme.

The necessity for a departure from classical mechanics is clearly shown by experimental results. In the first place the forces known in classical electrodynamics are inadequate for the explanation of the remarkable stability of atoms and molecules, which is necessary in order that materials may have any definite physical and chemical properties at all. The introduction of new hypothetical forces will not save the situation, since there exist general principles of classical mechanics, holding for all kinds of forces, leading to results in violent disagreement with observation. For example, if an atomic system has its equilibrium disturbed in any way and is then left alone, it will be set in oscillation and the oscillations will get impressed on the surrounding electromagnetic field, so that their frequencies may be observed with a spectroscope. Now whatever the laws of force governing the equilibrium, one would expect to be able to include the various frequencies in a scheme comprising certain fundamental frequencies and their harmonics. This is not observed to be the case. Instead, there is observed a new and unexpected connexion between the frequencies, embodied in Ritz's Combination Law of Spectroscopy, which is quite unintelligible from the classical standpoint.



One might try to get over the difficulty without departing from classical mechanics by assuming each of the spectroscopically observed frequencies to be a fundamental frequency with its own degree of freedom, the laws of force being such that the harmonic vibrations do not occur. Such a theory will not do, however, even apart from the fact that it would give no explanation of the Combination Law, since it would immediately bring one into conflict with the experimental evidence on specific heats. Classical statistical mechanics enables one to establish a general connexion between the total number of degrees of freedom of an assembly of vibrating systems and its specific heat. If one assumes all the spectroscopic frequencies of an atom to correspond to different degrees of freedom, one would get a specific heat for any kind of matter incomparably greater than the observed value. In fact the observed specific heats are given fairly well by a theory that takes into account merely the motion of each atom as a whole and assigns no internal motion to it at all.

This leads us to a new clash between classical mechanics and the results of experiment. There must certainly be some internal motion in an atom to account for its spectrum, but the internal degrees of freedom, for some classically inexplicable reason, do not contribute to the specific heat. A similar clash is found in connexion with the energy of oscillation of the electromagnetic field in a vacuum. Classical mechanics requires the specific heat corresponding to this energy to be infinite, but it is observed to be quite finite. A general conclusion from experimental results is that oscillations of high frequency do not contribute their classical quota to the specific heat.

As another illustration of the failure of classical mechanics we may consider the behaviour of light. We have, on the one hand, the phenomena of interference and diffraction, which can be explained only on the basis of a wave theory; on the other, phenomena such as photo-electric emission and scattering by free electrons, which show that light is composed of small particles. These particles, which are called photons, have each a definite energy and momentum, depending on the frequency of the light, and appear to have just as real an existence as electrons, or any other particles known in physics. A fraction of a photon is never observed.

Modern experiments have shown that this anomalous behaviour is not peculiar to light, but is quite general. All material particles have wave properties, which can be exhibited under suitable conditions. We



have here a very striking and general example of the break-down of classical mechanics—not merely an inaccuracy in its laws of motion, but *an inadequacy of its concepts to supply us with a description of atomic events.*

The necessity to depart from classical ideas when one wishes to account for the ultimate structure of matter may be seen, not only from experimentally established facts, but also from general philosophical grounds. In a classical explanation of the constitution of matter, one would assume it to be made up of a large number of small constituent parts and one would postulate laws for the behaviour of these parts, from which the laws of the matter in bulk could be deduced. This would not complete the explanation, however, since the question of the structure and stability of the constituent parts is left untouched. To go into this question, it becomes necessary to postulate that each constituent part is itself made up of smaller parts, in terms of which its behaviour is to be explained. There is clearly no end to this procedure, so that one can never arrive at the ultimate structure of matter on these lines. So long as *big* and *small* are merely relative concepts, it is no help to explain the big in terms of the small. It is therefore necessary to modify classical ideas in such a way as to give an absolute meaning to size.

At this stage it becomes important to remember that science is concerned only with observable things and that we can observe an object only by letting it interact with some outside influence. An act of observation is thus necessarily accompanied by some disturbance of the object observed. We may define an object to be big when the disturbance accompanying our observation of it may be neglected, and small when the disturbance cannot be neglected. This definition is in close agreement with the common meanings of big and small.

It is usually assumed that, by being careful, we may cut down the disturbance accompanying our observation to any desired extent. The concepts of big and small are then purely relative and refer to the gentleness of our means of observation as well as to the object being described. In order to give an absolute meaning to size, such as is required for any theory of the ultimate structure of matter, it becomes necessary to assume that *there is a limit to the fineness of our powers of observation and the smallness of the accompanying disturbance—a limit which is inherent in the nature of things and can never be surpassed by improved technique or increased skill on the part of the observer.* If the



object under observation is such that the limiting disturbance is negligible, then the object is big in the absolute sense and we may apply classical mechanics to it. If, on the other hand, the limiting disturbance is not negligible, then the object is small in the absolute sense and we require a new theory for dealing with it.

A consequence of the preceding discussion is that we must revise our ideas of causality. Causality applies only to a system which is left undisturbed. If a system is small, we cannot observe it without producing a serious disturbance and hence we cannot expect to find any causal connexion between the results of our observations. There is thus an essential indeterminacy in the quantum theory, of a kind that has no analogue in the classical theory, where causality reigns supreme. The quantum theory does not enable us in general to calculate the result of an observation, but only the probability of our obtaining a particular result when we make the observation.

The lack of determinacy in the quantum theory should not be considered as a thing to be regretted. It is necessary for a rational theory of the ultimate structure of matter. One of the most satisfactory features of the present quantum theory is that the differential equations that express the causality of classical mechanics do not get lost, but are all retained in symbolic form, and indeterminacy appears only in the application of these equations to the results of observations.

## 2. The Polarization of Photons

The discussion in the preceding section about the limit to the gentleness with which observations can be made and the consequent indeterminacy in the results of those observations does not provide any quantitative basis for the building up of quantum mechanics. For this purpose a new set of accurate laws of nature is required. One of the most fundamental and most drastic of these is the *Principle of Superposition of States*. We shall lead up to a general formulation of this principle through a consideration of some special cases, taking first the example provided by the polarization of light.

It is known experimentally that when plane-polarized light is used for ejecting photo-electrons, there is a preferential direction for the electron emission. Thus the polarization properties of light are closely connected with its corpuscular properties and one must ascribe a polarization to the photons. One must consider, for instance, a beam



of light plane-polarized in a certain direction as consisting of photons each of which is plane-polarized in that direction and a beam of circularly polarized light as consisting of photons each circularly polarized. Every photon is in a certain *state of polarization*, as we shall say. The problem we must now consider is how to fit in these ideas with the known facts about the resolution of light into polarized components and the recombination of these components.

Let us take a definite case. Suppose we have a beam of light passing through a crystal of tourmaline, which has the property of letting through only light plane-polarized perpendicular to its optic axis. Classical electrodynamics tells us what will happen for any given polarization of the incident beam. If this beam is polarized perpendicular to the optic axis, it will all go through the crystal; if parallel to the axis, none of it will go through; while if polarized at an angle  $\alpha$  to the axis, a fraction  $\sin^2\alpha$  will go through. How are we to understand these results on a photon basis?

A beam that is plane-polarized in a certain direction is to be pictured as made up of photons each plane-polarized in that direction. This picture leads to no difficulty in the cases when our incident beam is polarized perpendicular or parallel to the optic axis. We merely have to suppose that each photon polarized perpendicular to the axis passes unhindered and unchanged through the crystal, while each photon polarized parallel to the axis is stopped and absorbed. A difficulty arises, however, in the case of the obliquely polarized incident beam. Each of the incident photons is then obliquely polarized and it is not clear what will happen to such a photon when it reaches the tourmaline.

A question about what will happen to a particular photon under certain conditions is not really very precise. To make it precise one must imagine some experiment performed having a bearing on the question and inquire what will be the result of the experiment. Only questions about the results of experiments have a real significance and it is only such questions that theoretical physics has to consider.

In our present example the obvious experiment is to use an incident beam consisting of only a single photon and to observe what appears on the back side of the crystal. If one does this experiment, then according to quantum mechanics, sometimes one will find a whole photon, of energy equal to the energy of the incident photon, on the back side and other times one will find nothing. When one finds a



whole photon, it will be polarized perpendicular to the optic axis. One will never find only a part of a photon on the back side. If one does the experiment a large number of times, one will find the photon on the back side in a fraction  $\sin^2\alpha$  of the total number of times. Thus we may say that the photon has a probability  $\sin^2\alpha$  of passing through the tourmaline and appearing on the back side polarized perpendicular to the axis and a probability  $\cos^2\alpha$  of being absorbed. These values for the probabilities lead to the correct classical results for an incident beam containing a large number of photons.

In this way we preserve the individuality of the photon in all cases. We are able to do this, however, only because we abandon the determinacy of the classical theory. The result of an experiment is not determined, as it would be according to classical ideas, by the conditions under the control of the experimenter. The most that can be predicted is a set of possible results, with a probability of occurrence for each.

The foregoing discussion about the result of an experiment with a single obliquely polarized photon incident on a crystal of tourmaline answers all that can legitimately be asked about what happens to an obliquely polarized photon when it reaches the tourmaline. Questions about what decides whether the photon is to go through or not and how it changes its direction of polarization when it does go through cannot be investigated by experiment and should be regarded as outside the domain of science. Nevertheless some further description is necessary in order to correlate the results of this experiment with the results of other experiments that might be performed with photons and to fit them all into a general scheme. Such further description should be regarded, not as an attempt to answer questions outside the domain of science, but as an aid to the formulation of rules for expressing concisely the results of large numbers of experiments.

The further description provided by quantum mechanics runs as follows. It is supposed that a photon polarized obliquely to the optic axis may be regarded as being partly in the state of polarization parallel to the axis and partly in the state of polarization perpendicular to the axis. The state of oblique polarization may be considered as the result of some kind of superposition process applied to the two states of parallel and perpendicular polarization. This implies a certain special kind of relationship between the various states of



polarization, a relationship similar to that between polarized beams in classical optics, but which is now to be applied, not to beams, but to the states of polarization of one particular photon. This relationship allows any state of polarization to be resolved into, or expressed as a superposition of, any two mutually perpendicular states of polarization.

When we make the photon meet a tourmaline crystal, we are subjecting it to an observation. We are observing whether it is polarized parallel or perpendicular to the optic axis. The effect of making this observation is to force the photon entirely into the state of parallel or entirely into the state of perpendicular polarization. It has to make a sudden jump from being partly in each of these two states to being entirely in one or other of them. Which of the two states it will jump into cannot be predicted, but is governed only by probability laws. If it jumps into the parallel state it gets absorbed and if it jumps into the perpendicular state it passes through the crystal and appears on the other side preserving this state of polarization.

### 3. Interference of Photons

In this section we shall deal with another example of superposition. We shall again take photons, but shall be concerned with their position in space and their momentum instead of their polarization. If we are given a beam of roughly monochromatic light, then we know something about the location and momentum of the associated photons. We know that each of them is located somewhere in the region of space through which the beam is passing and has a momentum in the direction of the beam of magnitude given in terms of the frequency of the beam by Einstein's relation—momentum equals frequency multiplied by a universal constant. When we have such information about the location and momentum of a photon we shall say that it is in a definite *state of motion*. A state of motion is completely specified when one is given that it is associated with a certain beam.

We shall discuss the description which quantum mechanics provides of the interference of photons. Let us take a definite experiment demonstrating interference. Suppose we have a beam of light which is passed through some kind of interferometer, so that it gets split up into two components and the two components are subsequently made to interfere. We may, as in the preceding section, take



an incident beam consisting of only a single photon and inquire what will happen to it as it goes through the apparatus. This will present to us the difficulty of the conflict between the wave and corpuscular theories of light in an acute form.

Corresponding to the description that we had in the case of the polarization, we must now describe the photon as going partly into each of the two components into which the incident beam is split. The photon is then, as we may say, in a state of motion given by the superposition of the two states of motion associated with the two components. We are thus led to a generalization of the term 'state of motion' applied to a photon. For a photon to be in a definite state of motion it need not be associated with one single beam of light, but may be associated with two or more beams of light which are the components into which one original beam has been split.† In the accurate mathematical theory each state of motion is associated with one of the wave functions of ordinary wave optics, which wave function may describe either a single beam or two or more beams into which one original beam has been split. States of motion are thus superposable in a similar way to wave functions.

Let us consider now what happens when we determine the energy in one of the components. The result of such a determination must be either the whole photon or nothing at all. Thus the photon must change suddenly from being partly in one beam and partly in the other to being entirely in one of the beams. This sudden change is due to the disturbance in the state of motion of the photon which the observation necessarily makes. It is impossible to predict in which of the two beams the photon will be found. Only the probability of either result can be calculated from the previous distribution of the photon over the two beams.

One could carry out the energy measurement without destroying the component beam by, for example, reflecting the beam from a movable mirror and observing the recoil. Our description of the photon allows us to infer that, *after* such an energy measurement, it would not be possible to bring about any interference effects between the two components. So long as the photon is partly in one beam and partly in the other, interference can occur when the two beams are superposed,

† The circumstance that the superposition idea requires us to generalize our original meaning of states of motion, but that no corresponding generalization was needed for the states of polarization of the preceding section, is an accidental one with no underlying theoretical significance.



but this possibility disappears when the photon is forced entirely into one of the beams by an observation. The other beam then no longer enters into the description of the photon, which therefore counts as being entirely in the one beam in the ordinary way for any experiment that may subsequently be performed on it.

On these lines quantum mechanics is able to effect a union of the wave and corpuscular properties of light. The essential point is the association of each of the states of motion of a photon with one of the wave functions of ordinary wave optics. The nature of this association cannot be pictured on a basis of classical mechanics, but is something entirely new. It would be quite wrong to picture the photon and its associated wave as interacting in the way in which particles and waves can interact in classical mechanics. The association can be interpreted only statistically, the wave function giving us information about the probability of our finding the photon in any particular place when we make an observation of where it is.

Some time before the discovery of quantum mechanics people realized that the connexion between light waves and photons must be of a statistical character. What they did not clearly realize, however, was that the wave function gives information about the probability of *one* photon being in a particular place and not the probable number of photons in that place. The importance of the distinction can be made apparent in the following way. Suppose we have a beam of light consisting of a large number of photons split up into two components of equal intensity. On the assumption that the intensity of a beam is connected with the probable number of photons in it, we should have half the total number of photons going in each component. If the two components are now made to interfere, we should require a photon in one component to be able to interfere with one in the other. Sometimes these two photons would have to annihilate one another and other times they would have to produce four photons. This would contradict the conservation of energy. The new theory, which connects the wave function with probabilities for one photon, gets over the difficulty by making each photon go partly into each of the two components. Each photon then interferes only with itself. Interference between two different photons never occurs.

The association of particles with waves discussed above is not restricted to the case of light, but is, according to modern theory, of universal applicability. All kinds of particles are associated with



waves in this way and conversely all wave motion is associated with particles. Thus all particles can be made to exhibit interference effects and all wave motion has its energy in the form of quanta. The reason why these general phenomena are not more obvious is on account of a law of proportionality between the mass or energy of the particles and the frequency of the waves, the coefficient being such that for waves of familiar frequencies the associated quanta are extremely small, while for particles even as light as electrons the associated wave frequency is so high that it is not easy to demonstrate interference.

#### 4. Superposition and Indeterminacy

The reader may possibly be feeling dissatisfied with the attempt in the two preceding sections to fit in the existence of photons with the classical theory of light. He may argue that a very strange idea has been introduced—the possibility of a photon being partly in each of two states of polarization, or partly in each of two separate beams—but even with the help of this strange idea no satisfying picture of the fundamental single-photon processes has been given. He may say further that this strange idea did not provide any information about experimental results for the experiments discussed, beyond what could have been obtained from an elementary consideration of photons being guided in some vague way by waves. What, then, is the use of the strange idea?

In answer to the first criticism it may be remarked that the main object of physical science is not the provision of pictures, but is the formulation of laws governing phenomena and the application of these laws to the discovery of new phenomena. If a picture exists, so much the better; but whether a picture exists or not is a matter of only secondary importance. In the case of atomic phenomena no picture can be expected to exist in the usual sense of the word 'picture', by which is meant a model functioning essentially on classical lines. One may extend the meaning of the word 'picture' to include any *way of looking at the fundamental laws which makes their self-consistency obvious*. With this extension, one may acquire a picture of atomic phenomena by becoming familiar with the laws of the quantum theory.

With regard to the second criticism, it may be remarked that for many simple experiments with light, an elementary theory of waves



and photons connected in a vague statistical way would be adequate to account for the results. In the case of such experiments quantum mechanics has no further information to give. In the great majority of experiments, however, the conditions are too complex for an elementary theory of this kind to be applicable and some more elaborate scheme, such as is provided by quantum mechanics, is then needed. The method of description that quantum mechanics gives in the more complex cases is applicable also to the simple cases and although it is then not really necessary for accounting for the experimental results, its study in these simple cases is perhaps a suitable introduction to its study in the general case.

Before we can discuss the principle of superposition in the general case, we must introduce the important concept of a *state* of an atomic system. Let us take a general atomic system, composed of particles or bodies with specified properties (mass, moment of inertia, etc.) interacting according to specified laws of force. There will be various possible motions of the particles or bodies consistent with the laws of force. Each such motion is called a *state* of the system. According to classical ideas one could specify a state by giving numerical values to all the coordinates and velocities of the various component parts of the system at some instant of time, the whole motion being then completely determined. Now the argument of pages 3 and 4 shows that we cannot really observe a *small* system with that amount of detail which classical theory supposes. The limitation in the power of observation puts a limitation on the number of data that can be assigned to a state. Thus a state of an atomic system must be specified by fewer or more indefinite data than a complete set of numerical values for all the coordinates and velocities at some instant of time. In the case when the system is just a single photon, a state would be completely specified by a given state of motion in the sense of § 3 together with a given state of polarization in the sense of § 2.

A state of a system may be defined as a motion that is restricted by as many conditions or data as is possible without mutual disturbance or contradiction. In practice the conditions could be imposed by a suitable preparation of the system, consisting perhaps in passing it through various kinds of sorting apparatus, such as slits and polarimeters.

The general principle of superposition of quantum mechanics applies to the states, as thus defined, of any one dynamical system.



It requires us to assume that between these states there exist peculiar relationships such that whenever the system is definitely in one state we can consider it as being partly in each of two or more other states. The original state must be regarded as the result of a kind of *superposition* of the two or more new states, in a way that cannot be conceived on classical ideas. Any state may be considered as the result of a superposition of two or more other states, and indeed in an infinite number of ways. Conversely any two or more states may be superposed to give a new state. The procedure of expressing a state as the result of superposition of a number of other states is a mathematical procedure that is always permissible, independent of any reference to physical conditions, like the procedure of resolving a wave into Fourier components. Whether it is useful in any particular case, though, depends of course on the special physical conditions of the problem under consideration.

In the two preceding sections examples were given of the superposition principle applied to a system consisting of a single photon. § 2 dealt with states differing only with regard to the polarization and § 3 with states differing only with regard to the motion of the photon as a whole.

The nature of the relationships which the superposition principle requires to exist between the states of any system is of a kind that cannot be explained in terms of familiar physical concepts. One cannot in the classical sense picture a system being partly in each of two states and see the equivalence of this to the system being completely in some other state. There is an entirely new idea involved, to which one must get accustomed and in terms of which one must proceed to build up an exact mathematical theory, without having any detailed classical picture.

When a state is formed by the superposition of two other states, it will have properties that are in some vague way intermediate between those of the two original states and that approach more or less closely to those of either of them according to the greater or less 'weight' attached to this state in the superposition process. The new state is completely defined by the two original states when their relative weights in the superposition process are known, together with a certain phase difference, the exact meaning of weights and phases being provided in the general case by the mathematical theory. In the case of the polarization of a photon their meaning is that pro-



vided by classical optics, so that, for example, when two perpendicularly plane polarized states are superposed with equal weights, the new state may be circularly polarized in either direction, or linearly polarized at an angle  $\frac{1}{4}\pi$ , or else elliptically polarized, according to the phase difference.

The non-classical nature of the superposition process is brought out clearly if we consider the superposition of two states,  $A$  and  $B$ , such that there exists an observation which, when made on the system in state  $A$ , is certain to lead to one particular result,  $a$  say, and when made on the system in state  $B$  is certain to lead to some different result,  $b$  say. What will be the result of the observation when made on the system in the superposed state? The answer is that the result will be sometimes  $a$  and sometimes  $b$ , according to a probability law depending on the relative weights of  $A$  and  $B$  in the superposition process. It will never be different from both  $a$  and  $b$ . *The intermediate character of the state formed by superposition thus expresses itself through the probability of a particular result for an observation being intermediate between the corresponding probabilities for the original states,† not through the result itself being intermediate between the corresponding results for the original states.*

In this way we see that such a drastic departure from ordinary ideas as the assumption of superposition relationships between the states is possible only on account of the recognition of the importance of the disturbance accompanying an observation and of the consequent indeterminacy in the result of the observation. When an observation is made on any atomic system that is in a given state, in general the result will not be determinate, i.e., if the experiment is repeated several times under identical conditions several different results may be obtained. It is a law of nature, though, that if the experiment is repeated a large number of times, each particular result will be obtained in a definite fraction of the total number of times, so that there is a definite *probability* of its being obtained. This probability is what the theory sets out to calculate. Only in special cases when the probability for some result is unity is the result of the experiment determinate.

The assumption of superposition relationships between the states

† The probability of a particular result for the state formed by superposition is not necessarily intermediate between those for the original states in the general case when those for the original states are not zero or unity, so there are restrictions on the 'intermediateness' of a state formed by superposition.



leads to a mathematical theory in which the equations that define a state are linear in the unknowns. In consequence of this, people have tried to establish analogies with systems in classical mechanics, such as vibrating strings or membranes, which are governed by linear equations and for which, therefore, a superposition principle holds. Such analogies have led to the name 'Wave Mechanics' being sometimes given to quantum mechanics. It is important to remember, however, that *the superposition that occurs in quantum mechanics is of an essentially different nature from any occurring in the classical theory*, as is shown by the fact that the quantum superposition principle demands indeterminacy in the results of observations in order to be capable of a sensible physical interpretation. The analogies are thus liable to be misleading.

## 5. Mathematical Formulation of the Principle

Let us consider the whole set of states of a particular dynamical system. They will form an aggregate of things between which there will exist a number of relationships of a special kind, arising from the principle of superposition. These relationships we must now formulate in exact mathematical language.

The superposition process is a kind of additive process and implies that states can in some way be added to give new states. Now any mathematical quantities which can be added to give new quantities of the same nature may be represented by *vectors* in a suitable vector space with a sufficiently large number of dimensions. We are thus led to represent the states of a system by vectors in a certain vector space. The vectors will be assumed all to radiate from a common origin.

We represent each state by a vector denoted by a symbol  $\psi$ . Different states are represented by different vectors  $\psi$ , which may be distinguished by being provided with different suffixes; thus the states  $A, B, C$  may be represented by the vectors  $\psi_A, \psi_B, \psi_C$ . If the state  $A$  can be formed by superposition of the states  $B$  and  $C$ , then we assume that the corresponding vectors  $\psi_A, \psi_B, \psi_C$  are connected by an equation of the type

$$\psi_A = x_B \psi_B + x_C \psi_C, \quad (1)$$

where  $x_B$  and  $x_C$  are numbers.

From this assumption certain precise properties of the super-



position process follow—properties which are in fact necessary for the word ‘superposition’ to be suitable. Since, when vectors are added, the order in which they are put is unimportant, it follows that when two or more states are superposed, the order in which they occur in the superposition process is unimportant. The superposition process is symmetrical between the states that are superposed. Further, we see from equation (1) that (excluding the case when the coefficient  $x_B$  or  $x_C$  is zero) if the state  $A$  can be formed by superposition of the states  $B$  and  $C$ , then the state  $B$  can be formed by superposition of  $C$  and  $A$ , and  $C$  can be formed by superposition of  $A$  and  $B$ . The superposition relationship is symmetrical between all three states  $A$ ,  $B$ , and  $C$ . Three states that are symmetrically related in this way will be called *dependent*. More generally, any set of states  $A, B, \dots, Z$  will be called dependent if there exists a relation between their representative vectors of the form

$$x_A \psi_A + x_B \psi_B + \dots + x_Z \psi_Z = 0, \quad (2)$$

where the coefficients  $x_A, x_B, \dots, x_Z$  are not all zero; otherwise they will be called *independent*.

If we obtain the maximum number of independent states, this will give us the number of dimensions of our vector space. In most practical examples this number is infinite. The vector picture is useful in spite of this, most of the reasoning that we use it for being equally applicable whether the number of dimensions is finite or infinite.

To proceed with the accurate formulation of the superposition principle we must introduce a further assumption, namely the assumption that by superposing a state with itself we cannot form any new state, but only the original state over again. If the original state is represented by the vector  $\psi$ , when it is superposed with itself the result will be represented by

$$x_1 \psi + x_2 \psi = (x_1 + x_2) \psi,$$

where  $x_1$  and  $x_2$  are numbers. Now we may have  $x_1 + x_2 = 0$ , in which case the result of the superposition process would be nothing at all, the two components having cancelled each other by an interference effect. Our new assumption requires that, apart from this special case, the resulting state must be the same as the original one, so that  $(x_1 + x_2) \psi$  must represent the same state that  $\psi$  does. Now  $x_1 + x_2$  is an arbitrary number and hence we can conclude that *if the representative*



*vector of a state is multiplied by any number, not zero, the resulting vector will represent the same state.* Thus a state is specified by the direction of a vector in the vector space and any length one may assign to the vector is irrelevant. All the states of the dynamical system are in one-one correspondence with all the possible directions for a vector in the vector space, when one makes no distinction between the directions of the vectors  $\psi$  and  $-\psi$ .

The new assumption above shows up very clearly the fundamental difference between the superposition of the quantum theory and any kind of classical superposition. In the case of a classical system for which a superposition principle holds, for instance a vibrating membrane, when one superposes a state with itself the result is a *different* state, with a different magnitude of the oscillations. There is no physical characteristic of a quantum state corresponding to the magnitude of the classical oscillations, as distinct from their quality, described by the ratios of the amplitudes at different points of the membrane. Again, while there exists a classical state with zero amplitude of oscillation everywhere, namely the state of rest, there does not exist any corresponding state for a quantum system, the zero vector in the vector space representing no state at all.

One further assumption is necessary to complete the mathematical formulation of the principle of superposition. This is the assumption that in an equation expressing a superposition relationship, such as equation (1) or (2), the coefficients  $x$  can be complex numbers, and in the statement 'if the representative vector of a state is multiplied by any number, not zero, the resulting vector will represent the same state', the multiplying number can be complex.

The need for the allowing of complex coefficients can be seen in the two examples discussed in §§ 2 and 3, in which it is clear that from the superposition of two given states a *twofold* infinity of states may be obtained. In fact in the example of § 2, there are just two independent states of polarization for a photon, which may be taken to be the states of linear polarization parallel and perpendicular to some fixed direction, and from the superposition of these two a twofold infinity of states of polarization can be obtained, namely all the states of elliptic polarization, the general one of which requires two parameters to describe it. Again, in the example of § 3, from the superposition of two given states of motion for a photon a twofold infinity of states of motion may be obtained, the general one of which



is described by two parameters, which may be taken to be the ratio of the amplitudes of the two wave functions that are added together and their phase relationship. Now if, in the superposition equation (1), the coefficients  $x_B$ ,  $x_C$  were restricted to be real numbers, then, since only their ratio is of importance for determining the direction of the resultant vector  $\psi_A$  when  $\psi_B$  and  $\psi_C$  are given, there would be only a simple infinity of states obtainable from the superposition. The allowing of complex coefficients increases this to a twofold infinity.

Our assumption of complex coefficients implies that in every case of superposition of two different given states, a twofold infinity of states may be obtained. The vectors  $\psi$  representing the states are complex vectors, there being a twofold infinity of them with extremities on any given line in the vector space.

## 6. Analysis of the Principle

The principle of superposition that we have been discussing, applying to the states of any atomic system, is in agreement with the restricted principle of relativity, as it involves no reference to any particular Lorentz frame of reference. It would be desirable to develop the whole theory of quantum mechanics relativistically but at the present time this is not practicable, since relativistic quantum mechanics has as yet only a very limited applicability. There exists at present a general and logical scheme of non-relativistic quantum mechanics, yielding results in agreement with experiment, but, although one can obtain a formal extension of the scheme satisfying relativity requirements, this extension is not applicable to practical problems except with the help of approximations that are not mathematically justifiable.

The greater part of the present book will be concerned with the non-relativistic quantum mechanics, which is now as precise and as general as classical mechanics, to which it has, in fact, a strong analogy. The work will thus refer to one absolute time. The theory then naturally divides itself into two parts, part (i) dealing with relations and laws of nature governing the state of affairs in an atomic system at one instant of time, and part (ii) dealing with the connexion between the state of affairs at one instant of time and at a slightly later instant. Part (ii) will contain the analogue of the equations of motion of classical mechanics and will, in fact, be a neat



mathematical generalization of that scheme of equations. Part (i) will give essentially the theory of the limitations of one's powers of observation of a small system and its classical analogue will consist mainly of trivialities, since classical theory assumes there are no such limitations. A certain section of part (i), though (dealt with in Chapter V), will have a non-trivial classical analogue, concerned with the important dynamical notions of conjugate variables, contact transformations, and related things.

Historically, part (ii) was the first to be discovered. People guessed at the quantum generalization of the classical equations of motion and then proceeded to work with the quantum equations of motion, only gradually learning their proper physical significance and the limitations which they require in the possibilities of observation. In a logical exposition of the quantum theory, though, part (i) should be put first. This will accordingly be done in the present book, part (i) being dealt with in Chapters II to V and the equations of motion being then introduced in Chapter VI.

With the recognition of the natural separation of the theory into the two parts (i) and (ii), it becomes desirable to use the word 'state' in a rather different sense from that in which we have been using it up to the present. As we have been using it and as it comes in in the general formulation of the principle of superposition, it refers to the condition of the dynamical system throughout all time—something which, in the classical theory, would be described by a set of functions of the time which satisfy certain equations of motion. The preferable sense in which to use the word 'state' is to make it refer to the condition of the dynamical system at one instant of time—something which, in the classical theory, would be described merely by a set of numerical values for the dynamical variables. With the old meaning of the word, a dynamical system remains permanently in one state and just follows out the course of its motion in that state; with the new meaning a dynamical system is at each instant of time in a definite state and is continually changing from one state to another (or, as we may say, the state that the dynamical system is in is continually changing) under the influence of the equations of motion.

The old meaning is probably the more fundamental from an abstract theoretical point of view, since it is relativistic, referring to conditions throughout space-time, while the new meaning is non-relativistic, referring to conditions in a three-dimensional section of



space-time belonging to one time-instant. The new meaning, though, is better adapted to the line of development of the theory that we shall follow. It allows us to say that part (i) deals with the relations between the possible 'states' in which a dynamical system may be at any instant of time, and part (ii) deals with the connexion between the 'state' at one instant and that at a slightly later instant. The new meaning will therefore be used throughout the book,<sup>†</sup> except in a few places where otherwise stated.

If we now examine the general principle of superposition, applying to the 'states' of a system in the old sense, from our new non-relativistic point of view, we see that this principle resolves itself into two distinct hypotheses. One of these is a principle of superposition applying to the 'states' in the new sense. Between such states there must exist superposition relationships of just the same character as those between the old kind of states. The whole of § 5 will apply equally well to the new kind of states. The other hypothesis is that, if we take certain states at one instant of time that are connected by some superposition relationship, so that their representative vectors satisfy an equation of the type (1) or (2), then in the course of time these states will change in such a way that they always remain connected by this superposition relationship, their representative vectors  $\psi$  varying in such a way that equation (1) or (2) continues to hold, with constant coefficients. This second hypothesis turns the assumption of superposition relationships between the states at one instant of time into an assumption of superposition relationships holding between the various possible motions throughout all time, as required by the general principle of superposition.

The two hypotheses into which we have analysed the general principle of superposition belong respectively to parts (i) and (ii) of our theory. The principle of superposition of states in the new sense is one of the fundamental assumptions on which part (i) of the theory will be built, while the hypothesis of the constancy throughout time of any superposition relationship provides the basic assumption in the derivation of the equations of motion and the setting up of part (ii).

<sup>†</sup> This is an alteration from the first edition, where the old meaning was used throughout.



## II

### STATES AND OBSERVABLES

#### 7. The Vector Space representing the States

DURING the present century a profound change has taken place in the opinions physicists have held on the foundations of their subject. Previously they supposed that the principles of Newtonian mechanics would provide the basis for the description of the whole of physical phenomena and that all the theoretical physicist had to do was suitably to develop and apply these principles. With the recognition that there is no logical reason why Newtonian and other classical principles should be valid outside the domains in which they have been experimentally verified has come the modern point of view that departures from these principles are indeed necessary. Such departures find their expression through the introduction of new mathematical formalisms, new schemes of axioms and rules of manipulation, into the methods of theoretical physics.

Quantum mechanics provides a good example of the new ideas. It requires the states of a dynamical system and the observations that can be made on the system to be interconnected in ways that appear strange and unfamiliar from the classical standpoint. This results in the states and observations being represented by mathematical quantities of different natures from those ordinarily used. The new scheme becomes a precise physical theory when all the axioms and rules of manipulation governing the mathematical quantities are specified and when in addition certain laws are laid down connecting physical facts with the mathematical formalism, so that from any given physical conditions equations between the mathematical quantities may be inferred and vice versa. In an application of the theory one would be given certain physical information, which one would proceed to express by equations between the mathematical quantities. One would then deduce new equations with the help of the axioms and rules of manipulation and would conclude by interpreting these new equations as physical conditions. The justification for the whole scheme depends, apart from internal consistency, on the agreement of the final results with experiment.

The present chapter will be concerned with the foundation of the scheme in so far as it applies to the states of a dynamical system at



one particular time and observations made on the system at that time. We begin with the idea introduced in § 5 of representing each state by the direction of a vector in a certain vector space. (We saw in § 6 that this idea is valid for the states at a particular time.) It is now necessary to discuss the geometrical nature of the vector space—in particular the possibility of the existence of relations of perpendicularity between the vectors.

A convenient way of describing the geometrical nature of the vector space is by introducing a coordinate system of the simplest type possible and discussing the transformations of coordinates arising from the passage to other coordinate systems that are equally simple. Let the coordinates of a vector  $\psi_a$  be the set of numbers  $a_1, a_2, a_3, \dots$ . These numbers must in general be complex, since, as we saw in § 5, we can multiply the vectors by complex numerical coefficients and then add them to other vectors. If we make a passage to a new coordinate system, in which the coordinates of the vector  $\psi_a$  are  $a_1^*, a_2^*, a_3^*, \dots$ , then the new coordinates will be connected with the old ones by linear relations of the type

$$a_r^* = \sum_s \gamma_{rs} a_s, \quad (1)$$

where the  $\gamma_{rs}$  are numbers which depend only on the two coordinate systems and not on the vector  $\psi_a$ .

We now make the assumption that the  $\gamma_{rs}$  may be and in general are complex numbers, even when the two coordinate systems are both of the simplest type possible. The effect of this is that if the coordinates of  $\psi_a$  are real in one coordinate system they will in general be complex in the other. Thus one can give no invariant meaning to the vector  $\psi_a$  being real. *One cannot have a real vector in the vector space and one cannot split up a general vector into real and pure imaginary parts.*

Consider now the conjugate complex numbers to the coordinates of  $\psi_a$ . These conjugate complex numbers will also transform according to a linear law, namely the law

$$\bar{a}_r^* = \sum_s \bar{\gamma}_{rs} \bar{a}_s, \quad (2)$$

where the bar over a number denotes its conjugate complex. They may thus be considered as the coordinates of a vector in some vector space. It will be a different vector space from that of the  $\psi$ 's though, since the transformation law (2) is different from (1), on account of the



transformation coefficients  $\bar{\gamma}_{rs}$  being in general different from the  $\gamma_{rs}$ . There will be no meaning for the sum of one of the vectors in the new vector space with one of the  $\psi$ 's in the original vector space. The two vector spaces will not, of course, be entirely disconnected but must be related in a special way, since each transformation of coordinates in one of them is associated with a definite transformation of coordinates in the other.

We shall call the vectors in the new vector space  $\phi$ 's. That one of them whose coordinates are the conjugate complex numbers of the coordinates of a  $\psi$  with any specified suffix will be denoted by  $\phi$  with the same suffix. Thus  $\phi_a$  is the vector whose coordinates are the conjugate complex numbers to those of  $\psi_a$ . Two vectors such as  $\phi_a$  and  $\psi_a$ , whose coordinates are conjugate complex numbers, we shall define to be *conjugate imaginary* vectors. We use the words 'conjugate imaginary' instead of 'conjugate complex', since the relation between  $\phi_a$  and  $\psi_a$  is not quite the same as the relation between a pair of ordinary conjugate complex numbers, on account of its not being possible to add together  $\phi_a$  and  $\psi_a$  and to split up  $\phi_a$  and  $\psi_a$  into real and pure imaginary parts. The words 'conjugate complex' and the notation of putting a bar over a quantity to get its conjugate complex will be reserved for quantities which can be split up into real and pure imaginary parts. Thus we shall speak of the conjugate imaginary of a vector  $\phi$ , but of the conjugate complex of the coordinates of this vector in any specified coordinate system, the coordinates being just ordinary numbers.

Each vector  $\psi_a$  in the space of  $\psi$ 's determines uniquely a vector  $\phi_a$  in the space of  $\phi$ 's and vice versa. Thus the space of  $\phi$ 's provides a representation of the states of our dynamical system just as well as the space of  $\psi$ 's, each state being associated with one direction in the space of  $\phi$ 's. There is, in fact, perfect symmetry between the  $\phi$ 's and  $\psi$ 's, which symmetry will survive all through the theory.

We now introduce a further and final geometrical property of the space of  $\psi$ 's. We assume that, if  $a_1, a_2, a_3, \dots$  and  $b_1, b_2, b_3, \dots$  are the coordinates of any two vectors  $\psi_a$  and  $\psi_b$  referred to one of the simplest coordinate systems, then, in the passage to any other of the simplest coordinate systems, the coordinates will transform in such a way that the number

$$\bar{a}_1 b_1 + \bar{a}_2 b_2 + \bar{a}_3 b_3 + \dots \quad (3)$$

remains invariant. This assumption imposes certain conditions on



the coefficients  $\gamma_{rs}$  in (1). The number (3) may be regarded as the scalar product of the vector  $\psi_b$  with the vector  $\phi_a$ , the conjugate imaginary of  $\psi_a$ , and may be denoted by the symbolic product  $\phi_a \psi_b$ . There is no invariant of the type  $a_1 b_1 + a_2 b_2 + a_3 b_3 + \dots$  or  $\bar{a}_1 \bar{b}_1 + \bar{a}_2 \bar{b}_2 + \bar{a}_3 \bar{b}_3 + \dots$ , corresponding to the scalar product of  $\psi_a$  with  $\psi_b$  or of  $\phi_a$  with  $\phi_b$ , and thus symbolic products of the type  $\psi_a \psi_b$  or  $\phi_a \phi_b$  never occur in the theory.

An invariant of the type (3) is not unusual in pure mathematics. It forms an interesting generalization to the case of complex coordinates of the ordinary scalar product  $a_1 b_1 + a_2 b_2 + a_3 b_3 + \dots$  of two vectors with real coordinates  $a_1, a_2, a_3, \dots$  and  $b_1, b_2, b_3, \dots$  in ordinary Euclidean space. The invariant (3) introduces a closer and more familiar connexion between the  $\psi$ 's and the  $\phi$ 's. Instead of picturing the  $\psi$ 's and  $\phi$ 's as vectors in two different vector spaces, we may picture them as two different kinds of vector associated with the same space. The relation between these two kinds of vector is then just the one well known in differential geometry as the relation between covariant and contravariant vectors.

The symbolic product notation  $\phi_a \psi_b$  is very convenient for general discussions and will be extensively used in this book. When using it, we shall make the convention always to put the  $\phi$ -symbol to the left of the  $\psi$ -symbol, since the notation then fits in very well with the matrix notation that will be developed later. As before remarked, products like  $\psi_a \psi_b$  and  $\phi_a \phi_b$  never occur.

From the definition (3) it follows at once that the symbolic product  $\phi_a \psi_b$  is subject to the usual algebraic axioms for the product of two quantities, as exemplified by the following equations:

$$\phi_a(\psi_b + \psi_c) = \phi_a \psi_b + \phi_a \psi_c,$$

$$(\phi_a + \phi_b)\psi_c = \phi_a \psi_c + \phi_b \psi_c,$$

and

$$\phi_a(k\psi_b) = (k\phi_a)\psi_b = k(\phi_a \psi_b)$$

where  $k$  is any number. Further results that follow immediately from the definition are that the two numbers  $\phi_a \psi_b$  and  $\phi_b \psi_a$  are conjugate complex, i.e.

$$\phi_a \psi_b = \overline{\phi_b \psi_a}, \quad (4)$$

and the number  $\phi_a \psi_a$  is always real and positive except in the special case when the vector  $\psi_a$  vanishes. This number  $\phi_a \psi_a$  may be called the square of the length of the vector  $\psi_a$  or of the vector  $\phi_a$ , in agreement with the meaning of length for an ordinary vector with real coordinates.



It will frequently happen in the course of development of the theory and also in its applications that we shall have to introduce a vector  $\psi$  or  $\phi$  whose direction is fixed by special considerations referring to the problem in hand, but whose length is not so fixed. It is then often convenient to choose the length to be equal to unity. This procedure is called *normalization* and the vector so chosen that its length is unity is said to be *normalized*. It should be noted that the vector is not, even then, completely determined, since one can always multiply it by any number of modulus unity, i.e. of the type  $e^{ic}$  where  $c$  is real, without changing either its direction or its length. We call such a number a *phase factor*.

If a  $\phi$ -vector and a  $\psi$ -vector are such that their product  $\phi\psi$  is zero, we shall say that these two vectors are *orthogonal*. We shall also say that two  $\psi$ 's are orthogonal if the product of either with the conjugate imaginary of the other is zero. Thus  $\psi_a$  and  $\psi_b$  are orthogonal if  $\phi_a\psi_b = 0$  or if  $\phi_b\psi_a = 0$ , these two conditions being, of course, equivalent on account of (4). A similar definition will hold for the orthogonality of two  $\phi$ 's. Further, we shall say that two states are orthogonal if the vectors representing these states are orthogonal.

## 8. Observables as Linear Operators

The preceding section completes all that can be said about the relationships between the states of a system at a particular time. To continue with the development of the theory we must introduce observations into the discussion. We shall be concerned here only with observations made at this same particular time. Each such observation consists in the measurement of the value at this time of some dynamical coordinate or momentum, or some function of the coordinates and momenta. It will be convenient to introduce a special word for these things that get measured, as they play such an important part in the theory. We shall call each of them an *observable*. Thus an observation consists in the measurement of an observable.

In the present section we shall deal only with the general relations which exist between observables and which connect observables with states. The discussion of the measurements of observables and the way in which the numerical results of such measurements appear in the theory will be left to the next section.

We make the fundamental assumption that *each observable is*



represented in the mathematical formalism by a linear operator that can operate on the  $\psi$ -vectors. By a linear operator is meant an operator which, operating on any  $\psi$ -vector, changes that  $\psi$ -vector into another  $\psi$ -vector whose coordinates are linear functions of the coordinates of the first one. Thus, when it operates on the vector  $\psi_x$  with coordinates  $x_r$  ( $r = 1, 2, 3, \dots$ ), it will change that vector into some vector,  $\psi_b$  say, whose coordinates  $b_r$  are connected with the  $x_r$  by relations of the type

$$b_r = \sum_s \alpha_{rs} x_s, \quad (5)$$

where the  $\alpha_{rs}$  are numbers (in general complex), which depend only on the operator and not on the vector  $\psi_x$ .

The numbers  $\alpha_{rs}$  may be called the coordinates of the operator. They differ from the coordinates of a vector in that there are many more of them, each of them requiring two suffixes to label it instead of one. If we had to write them out explicitly, the natural way of arranging them would be as a two-dimensional array, thus:

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & . & . & . \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & . & . & . \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \end{bmatrix}.$$

Such an array is called a *matrix* and the numbers are called the *elements* of the matrix. We make the convention that the elements must always be arranged so that those in the same row have the same first suffix and those in the same column have the same second suffix. An element whose two suffixes are the same, such as  $\alpha_{rr}$ , is called a *diagonal element*, as all such elements lie on a diagonal of the array.

There is a symbolic notation which can conveniently be used in connexion with linear operators, corresponding to that which we had for the scalar product in the preceding section. The linear operator with coordinates  $\alpha_{rs}$  we call the linear operator  $\alpha$ , and when it acts on any vector  $\psi_x$ , the resulting vector is regarded as the symbolic product of  $\alpha$  with  $\psi_x$  and is written  $\alpha\psi_x$ , with the operator to the left of the  $\psi$ -symbol. Thus when the coordinates of two vectors  $\psi_x$  and  $\psi_b$  and of a linear operator  $\alpha$  are connected by equation (5), the relation between the vectors and linear operator may be expressed by the symbolic equation

$$\psi_b = \alpha\psi_x. \quad (6)$$



It is easily seen that for any two vectors  $\psi_x$  and  $\psi_y$ ,

$$\alpha(\psi_x + \psi_y) = \alpha\psi_x + \alpha\psi_y,$$

and for any number  $k$ ,

$$\alpha(k\psi_x) = k(\alpha\psi_x).$$

These equations, in fact, are just those that express in the symbolic notation the linearity property of the operator  $\alpha$ .

The assumption that observables are represented by linear operators seems at first sight to be a very drastic and unexpected departure from ordinary ideas. It appears much more reasonable, though, when one examines the properties of linear operators and sees how well fitted they are to play the part of observables. In the mathematics of classical mechanics, observables are quantities which we can add to one another or multiply with one another or form algebraic functions of, the results of these processes being other observables. Now the theory of linear operators can be developed so that we can add and multiply linear operators and form algebraic functions of them, the results of these processes being other linear operators. Thus linear operators can be handled mathematically in much the same way in which one is used to handling observables in classical mechanics.

The sum of two linear operators is defined as that linear operator which, operating on any vector  $\psi_x$ , changes this vector into the sum of the two vectors into which  $\psi_x$  is changed by the two operators individually. Thus, in the symbolic notation, the sum  $\alpha + \beta$  of two linear operators  $\alpha$  and  $\beta$  is defined by the equation

$$(\alpha + \beta)\psi_x = \alpha\psi_x + \beta\psi_x$$

holding for all  $\psi_x$ . Similarly, the product of a linear operator  $\alpha$  with a number  $k$  is defined as that linear operator which, operating on any vector  $\psi_x$ , changes this vector into  $k$  times the vector into which  $\psi_x$  is changed by  $\alpha$ . In the symbolic notation,

$$(k\alpha)\psi_x = k(\alpha\psi_x).$$

With the help of these two definitions one can form linear functions, with arbitrary numerical coefficients, of the linear operators.

The product of two linear operators is defined as that linear operator which produces by itself the same effect as the two operators in succession. Thus the product  $\alpha\beta$  is defined as the operator which, operating on a vector  $\psi_x$ , changes it into that vector which one would get by operating first on  $\psi_x$  with  $\beta$ , and then on the result of the first



operation with  $\alpha$ . In symbols,

$$(\alpha\beta)\psi_x = \alpha(\beta\psi_x).$$

In general this would not be the same as operating first with  $\alpha$  and then with  $\beta$ , so that the product  $\alpha\beta$  in general differs from  $\beta\alpha$ . *The commutative axiom of multiplication does not hold for linear operators.* It may happen in a special case that two linear operators  $\alpha$  and  $\beta$  are such that  $\alpha\beta$  and  $\beta\alpha$  are equal. In this case we say that  $\alpha$  *commutes* with  $\beta$ , or that  $\alpha$  and  $\beta$  *commute*.

It is easily seen that the other multiplication axioms of ordinary algebra, the associative and distributive axioms, as well as all the addition axioms, are valid for linear operators. Thus one can build up an algebra for linear operators which runs very similar to ordinary algebra. For instance, by repeated applications of the processes of addition and multiplication one can construct functions of linear operators, in fact all those functions that can be expressed as power series.

When we say that observables are *represented* by linear operators, it is implied that the algebraic relations which exist between any observables are the same as the algebraic relations between the linear operators representing those observables. Thus the essential mathematical significance of the assumption that observables are represented by linear operators is that *observables are subject to an algebra which is the same as ordinary algebra with the exception that the commutative axiom of multiplication does not hold.*

Up to the present we have considered observables only in connexion with  $\psi$ -vectors. To maintain symmetry between the  $\psi$ 's and the  $\phi$ 's, we must have the possibility of representing an observable by a linear operator operating on the  $\phi$ -vectors. This possibility can be deduced from the assumption that observables are represented by linear operators operating on the  $\psi$ -vectors, together with the relations between  $\phi$ 's and  $\psi$ 's established in the preceding section.

Let us form the scalar product of the  $\psi$ -vector  $\alpha\psi_x$  with an arbitrary  $\phi$ -vector  $\phi_y$ . If the coordinates of  $\phi_y$  are the numbers  $\bar{y}_r$ , then, the coordinates of  $\alpha\psi_x$  being given by the right-hand side of (5), the scalar product has the value

$$\phi_y(\alpha\psi_x) = \sum_r \bar{y}_r \sum_s \alpha_{rs} x_s.$$



This may be written in the form

$$\sum_s \bar{a}_s x_s, \quad (7)$$

where

$$\bar{a}_s = \sum_r \bar{y}_r \alpha_{rs}. \quad (8)$$

Now the  $\bar{a}_s$  must be such that expression (7) is invariant under a change of coordinate system when the  $x_s$  are the coordinates of any  $\psi$ -vector. This is sufficient to ensure that the  $\bar{a}_s$  transform like the coordinates of a  $\phi$ -vector and thus that the  $\bar{a}_s$  are the coordinates of some  $\phi$ -vector. We call this vector  $\phi_y \alpha$  and consider it as the symbolic product of the vector  $\phi_y$  with an operator  $\alpha$ . It is, in fact, the result of some linear operator operating on  $\phi_y$ , since, as shown by equation (8), its coordinates are linear functions of the coordinates of  $\phi_y$ . This linear operator provides an alternative representation for the observable symmetrical with the previous linear operator operating on  $\psi$ -vectors.

The use of the same letter  $\alpha$  for both operators makes a convenient notation which does not lead to confusion. In fact, the two operators may be counted as just one operator which can operate either to the right on a  $\psi$ -vector or to the left on a  $\phi$ -vector. We now have a symbolic scheme in which the following associative law of multiplication holds,

$$\phi_y (\alpha \psi_x) = (\phi_y \alpha) \psi_x.$$

Either of these quantities will be written in future as  $\phi_y \alpha \psi_x$  without brackets.

A number, considered as a multiplying factor into each  $\psi$  or each  $\phi$ , is a special case of a linear operator. It has the property of commuting with every linear operator. One can easily see that any linear operator that commutes with every linear operator is a number.

One further question remains to be considered in this section. We are assuming that every observable can be represented by a linear operator. Does every linear operator represent some observable? One would immediately expect some restriction on the linear operator of the nature of a condition of reality, since, owing to the fact that a linear operator may be multiplied by an arbitrary complex number and remains a linear operator, the general linear operator must correspond to a complex function of the dynamical variables. Such a complex function may, of course, be considered formally as a complex observable, but since no meaning can be attached to the measurement



of a complex observable,† it is preferable to restrict the word ‘observable’ to refer to real functions of dynamical variables and to introduce a corresponding restriction on the linear operators that represent observables.

We assume this restriction to be that the coordinates  $\alpha_{rs}$  of each of these linear operators satisfy the relations

$$\alpha_{rs} = \overline{\alpha_{sr}}. \tag{9}$$

These are just the relations required for the matrix formed by the  $\alpha_{rs}$  to be what is called a *Hermitian matrix*. A linear operator that satisfies this condition may conveniently be called a *Hermitian operator*.

In order to see that this assumption is suitable, we must verify that the condition imposed by (9) is independent of the system of coordinates. An easy way of doing this is by putting this condition in a form that does not refer to any coordinate system. Introducing two arbitrary  $\psi$ -vectors,  $\psi_x$  and  $\psi_y$  say, with coordinates  $x_r$  and  $y_r$ , we have

$$\phi_x \alpha \psi_y = \sum_{rs} \bar{x}_r \alpha_{rs} y_s$$

and

$$\phi_y \alpha \psi_x = \sum_{rs} \bar{y}_r \alpha_{rs} x_s = \sum_{rs} x_r \alpha_{sr} \bar{y}_s.$$

From (9) we now obtain

$$\phi_x \alpha \psi_y = \overline{\phi_y \alpha \psi_x}. \tag{10}$$

Conversely, if we are given (10) for arbitrary  $\psi_x$  and  $\psi_y$ , we can deduce (9), by taking  $\psi_x$  and  $\psi_y$  to be the unit vectors along the directions of the  $r$ -th and  $s$ -th axes respectively. Thus the condition on  $\alpha$  imposed by (9) is equivalent to that imposed by (10), and since (10) contains no reference to any coordinate system (9) must also be independent of the coordinate system.

As a corollary to this work we see, by putting  $\alpha \psi_x = \psi_b$  in (10),

$$\phi_x \alpha \psi_y = \overline{\phi_y \psi_b} = \phi_b \psi_y,$$

from (4). Since this holds for arbitrary  $\psi_y$ , we must have

$$\phi_x \alpha = \phi_b. \tag{11}$$

Thus the  $\psi$ -vector  $\alpha \psi_x$  has as its conjugate imaginary the  $\phi$ -vector  $\phi_x \alpha$ . This result will be much used in the following work. It is true only on account of the operator  $\alpha$  being Hermitian.

† It would not do to measure separately the real and pure imaginary parts, because this would mean two measurements, which in general would interfere with one another.



The question now remains—Does every Hermitian operator represent an observable? The answer to this is that, provided we give a sufficiently comprehensive meaning to the word observable, to make it include all real functions of the dynamical variables that are theoretically measurable and not merely those for which a practicable method of measurement can be set up, most of the Hermitian operators ordinarily met with do represent observables, but there are exceptions. The remaining condition for a Hermitian operator to represent an observable will be given at the end of § 10.

## 9. Eigenvalues

In the two preceding sections we made a number of assumptions about the way in which states and observables are to be represented mathematically in the theory. These assumptions are not, by themselves, laws of nature, but become laws of nature when we make some further assumptions that provide a physical interpretation of the theory. Such further assumptions must take the form of establishing connexions between physical facts, on the one hand, and the equations of the mathematical formalism on the other.

One of these further assumptions is the following: *In the special case when the result of a particular observation made on the system in a particular state is with certainty one particular number,  $a$  say (instead of being one of two or more numbers according to a probability law), then the Hermitian operator,  $\alpha$  say, representing the observable that is measured and the  $\psi$ -vector,  $\psi_a$  say, representing the state are connected by the equation*

$$\alpha\psi_a = a\psi_a. \quad (12)$$

*Conversely, if this equation holds, a measurement of the observable represented by  $\alpha$  made on the system in the state represented by  $\psi_a$  is certain to lead to the result  $a$ .*

Equation (12) means that the linear operator  $\alpha$ , applied to the vector  $\psi_a$ , just multiplies this vector by a numerical factor without changing its direction (or possibly multiplies it by the factor zero, so that it ceases to have a definite direction). This same  $\alpha$  applied to other vectors will, of course, in general change both their lengths and their directions. It should be noticed that only the direction of  $\psi_a$  is of importance in equation (12). If one multiplies  $\psi_a$  by any number not zero, it will not affect the question of whether  $\psi_a$  satisfies equation (12) or not. This, of course, is necessary in order that our



assumption may be sensible, since the state represented by  $\psi_a$  depends only on the direction of  $\psi_a$  and not on its length.

There are some other matters which we must look into before we can be sure that our assumption is reasonable. One of these concerns the reality of the number  $a$ . Any result of a measurement is necessarily a real number. Is any number  $a$  satisfying an equation of the type (12) also necessarily real? We can easily see that it is so when we make use of the Hermitian property of  $\alpha$ . Multiplying (12) symbolically by  $\phi_a$ , the conjugate imaginary of  $\psi_a$ , on the left, we get

$$\phi_a \alpha \psi_a = a \phi_a \psi_a.$$

Now  $\phi_a \alpha \psi_a$  is a real number, as follows from equation (10) with  $\psi_x$  and  $\psi_y$  both put equal to  $\psi_a$ , and further we saw, just after equation (4), that  $\phi_a \psi_a$  must be real. Hence  $a$  must be a real number.

Another point to be noticed is that our assumption does not disturb the symmetry between  $\psi$ 's and  $\phi$ 's. We can, in fact, replace equation (12) in the assumption by

$$\phi_a \alpha = a \phi_a, \quad (13)$$

equations (12) and (13) being equivalent since they are just conjugate imaginary equations, according to the rule deduced in the preceding section in connexion with equation (11).

A further question to be looked into is the following. If we have any observable, we can multiply it by any real number  $k$  and get another observable. Now if a measurement of the original observable with the system in a particular state is certain to lead to one particular result  $a$ , we should require for physical consistency that a measurement of the new observable with the system in this same state shall certainly lead to the result  $ka$ . Is this given by the mathematical formalism, with the help of our assumption? It is easily seen that it is. If the original observable is represented by the operator  $\alpha$ , the new one must be represented by  $k\alpha$ . The condition that a measurement of the original observable shall certainly lead to the result  $a$  when the system is in the state represented by  $\psi_a$  is equation (12) and from this equation we can deduce

$$(k\alpha)\psi_a = ka\psi_a,$$

from which we can infer that a measurement of the new observable will certainly lead to the result  $ka$  for this same state.

The above question is a special case of a more general one. We may take as new observable any function  $f$  of the original one, instead of just  $k$  times it, and we should then require for physical consistency



that a measurement of the new observable shall certainly lead to the result  $f(a)$ . This also is deducible from the mathematical formalism in an elementary way, provided the function  $f$  is expressible as a power series. The general case, where the function  $f$  may not be expressible as a power series, will be dealt with in § 11, where the requirement we are now discussing will be used as a basis for a general mathematical definition of a function of an observable.

Equation (12) is of such fundamental importance in the theory that it is desirable to introduce some special words to describe the relationships between the quantities involved. We shall call  $a$  an *eigenvalue*<sup>†</sup> of the operator  $\alpha$  or of the observable that  $\alpha$  represents and  $\psi_a$  an *eigen- $\psi$*  of this operator or observable, and we shall say that the eigen- $\psi$   $\psi_a$  belongs to the eigenvalue  $a$ . Likewise,  $\phi_a$  satisfying (13) is an *eigen- $\phi$*  belonging to the eigenvalue  $a$ , and the state represented by either  $\psi_a$  or  $\phi_a$  is an *eigenstate* belonging to this eigenvalue. This terminology may also be used when the linear operator  $\alpha$  is not Hermitian and does not represent an observable.

Our assumption now enables us to infer that *every eigenvalue of an observable is a possible result of the measurement of that observable*. It is certainly the result when the system is in an eigenstate belonging to the eigenvalue. The converse theorem, that *every possible result of the measurement of an observable, with the system in any state whatever, is one of its eigenvalues*, is also true and will be deduced in § 12 from a more general assumption for physical interpretation. The set of eigenvalues of an observable are just the possible results of measurements of that observable and the calculation of eigenvalues is thus an important practical problem.

A real number  $k$  is a special case of a Hermitian operator. Its peculiar characteristic from our present standpoint is that it has just one eigenvalue, namely  $k$ , and every  $\psi$ -vector is an eigen- $\psi$  belonging to this eigenvalue. Such an operator may be considered to represent an observable. Any measurement of the observable must then always lead to the same result, namely  $k$ , no matter what state the system is in. This means that the observable is a natural constant, such as the velocity of light or the charge of an electron, or perhaps just a number.

<sup>†</sup> The word 'proper' is sometimes used instead of 'eigen', but this is not satisfactory as the words 'proper' and 'improper' are so often used with other meanings. See, for example, § 20 and also p. 181.



The theorem will now be proved that *two eigenstates belonging to two different eigenvalues of an observable are orthogonal*, in the sense defined at the end of § 7. Suppose the two eigenstates are represented by the eigen- $\psi$ 's  $\psi_1$  and  $\psi_2$ , the corresponding eigenvalues being  $a_1$  and  $a_2$  respectively. Then, if  $\alpha$  represents the observable, we have

$$\alpha\psi_1 = a_1\psi_1 \quad (14)$$

and 
$$\phi_2\alpha = a_2\phi_2, \quad (15)$$

the second of these equations being of the type (13). Multiplying (14) by  $\phi_2$  on the left-hand side and (15) by  $\psi_1$  on the right-hand side, we obtain

$$\phi_2\alpha\psi_1 = a_1\phi_2\psi_1$$

and 
$$\phi_2\alpha\psi_1 = a_2\phi_2\psi_1$$

respectively. Hence 
$$a_1\phi_2\psi_1 = a_2\phi_2\psi_1,$$

so that if  $a_1$  is not equal to  $a_2$ ,  $\phi_2\psi_1 = 0$  and the two states are orthogonal. We shall call this theorem the *orthogonality theorem*.

If  $\psi_1$  and  $\psi_2$  are two eigen- $\psi$ 's belonging to the same eigenvalue, then it is easily seen that any linear combination of them,  $c_1\psi_1 + c_2\psi_2$ , is also an eigen- $\psi$  belonging to this eigenvalue. Physically, this means that if we take two states such that a measurement of some observable with the system in either of these states is certain to lead to one particular result, then a measurement of this observable with the system in any state formed by the superposition of the two states will also certainly lead to this same result. This gives us some understanding of the physical meaning of superposition.

It can easily be proved that no linear combination of eigen- $\psi$ 's belonging to different eigenvalues can be an eigen- $\psi$ , i.e., that *eigen- $\psi$ 's belonging to different eigenvalues are all necessarily independent*. If this were not so, we should have a relation of the type

$$\sum_r c_r \psi_r = 0, \quad (16)$$

with numerical coefficients  $c_r$ , between a number of eigen- $\psi$ 's belonging to different eigenvalues. In this relation we can assume, without loss of generality, that there is only one term corresponding to any eigenvalue, since if there were several terms corresponding to the same eigenvalue, these terms could be lumped together to form a single term, which would still be an eigen- $\psi$ . If we now multiply (16) on the left by  $\bar{c}_s\phi_s$ , the conjugate imaginary of one of the terms, we get, on account of the orthogonality theorem proved above,

$$\bar{c}_s\phi_s c_s \psi_s = 0,$$







This is an algebraic equation of the  $n$ -th degree in  $a$  and must have  $n$  roots, not necessarily all different. Each of these roots is an eigenvalue and the eigen- $\psi$  belonging to it may be obtained from (18). When two or more of the roots of (19) coincide at some particular value,  $a_1$  say, then the eigenvalue  $a_1$  must have a number of independent eigen- $\psi$ 's belonging to it equal to the number of these coincident roots. This result may be proved by algebraic methods, but one can also see that it must be true from elementary considerations of continuity. Suppose small variations, of the order of magnitude of  $\epsilon$ , to be made in the matrix elements  $\alpha_{rs}$  in such a way as not to destroy the Hermitian character of the matrix and so as to separate all the roots,  $m$  in number, say, that previously coincided at  $a_1$ . These roots will then differ from one another and from  $a_1$  by quantities of the order of  $\epsilon$ . Each of them will have some eigen- $\psi$  belonging to it and these eigen- $\psi$ 's will be all orthogonal to one another, by the orthogonality theorem of the preceding section. These eigen- $\psi$ 's will define an  $m$ -dimensional sub-space, which contains them all. Any  $\psi$  in this sub-space (with a length of the order unity) will satisfy the condition for being an eigen- $\psi$  of the original Hermitian operator  $\alpha$  belonging to the eigenvalue  $a_1$ , with an error of the order  $\epsilon$ . By now making  $\epsilon \rightarrow 0$ , we obtain in the limit an  $m$ -dimensional sub-space of  $\psi$ 's and thus  $m$  independent  $\psi$ 's, each of which is an eigen- $\psi$  of  $\alpha$  belonging to the eigenvalue  $a_1$ .

It should be noted that the argument makes use of the Hermitian property of  $\alpha$  in using the orthogonality theorem and nowhere else. It is necessary to use this theorem, since otherwise some of the eigen- $\psi$ 's of the varied Hermitian operator, belonging to eigenvalues near  $a_1$ , might be inclined to each other at angles of the order  $\epsilon$  and tend to coincidence as  $\epsilon \rightarrow 0$ , in which case the argument would fail. A simple example of such failure is obtained if we take the non-Hermitian matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

We now have the result that, when the number of dimensions of the  $\psi$ -space is finite and equal to  $n$ , the number of independent eigen- $\psi$ 's of any Hermitian operator is also  $n$ . Hence an arbitrary  $\psi$  can be expressed linearly in terms of these eigen- $\psi$ 's, thus

$$\psi = \sum \psi_a, \quad (20)$$

where  $\psi$  is arbitrary and each  $\psi_a$  is an eigen- $\psi$ . This is the expansion



theorem for the case when the number of dimensions of the  $\psi$ -space is finite.

We must now go over to the case of an infinite number of dimensions. The expansion theorem still reads—*an arbitrary  $\psi$  can be expressed linearly in terms of the eigen- $\psi$ 's of a Hermitian operator*, but the theorem can no longer always be written in the form (20), since the number of independent eigen- $\psi$ 's may be more than enumerable. This happens when the eigenvalues include all numbers in a certain range, say all numbers from  $p$  to  $q$ . (It may quite possibly be from  $-\infty$  to  $+\infty$ .) The expansion may then take the form of an integral,

$$\psi = \int_p^q \psi_a da, \quad (21)$$

where  $a$  is the eigenvalue to which  $\psi_a$  belongs and  $\psi_a$  varies with  $a$  in such a way that the integral exists. This form of expansion, though, does not include all cases. We shall take up this question again in § 20.

A rigorous proof of the expansion theorem, sufficiently general to cover all the cases for which it is required in quantum theory, has not yet been found. The following argument, however, makes the theorem appear plausible.

Let  $\alpha$  be the Hermitian operator and consider a  $\psi$ -vector  $\psi_\tau$  that is a function of the parameter  $\tau$  and satisfies the differential equation

$$\frac{d}{d\tau} \psi_\tau = i\alpha \psi_\tau. \quad (22)$$

If  $\psi_\tau$  is given for one value of  $\tau$ , then it is fixed by this equation for a slightly greater value of  $\tau$ . Thus we should expect this equation to have one solution, and only one, for any given initial value for  $\psi_\tau$ , i.e., for  $\psi_\tau$  equal to an arbitrary  $\psi_0$  when  $\tau = 0$ . Suppose now that this solution can be expressed as a Fourier series or integral in  $\tau$ ; thus, if we take for definiteness the case of the integral,

$$\psi_\tau = \int e^{ia\tau} \psi_a da, \quad (23)$$

where  $\psi_a$  is independent of  $\tau$ , but involves the new parameter  $a$ . Substituting this expression for  $\psi_\tau$  in (22), we obtain

$$\int iae^{ia\tau} \psi_a da = i\alpha \int e^{ia\tau} \psi_a da,$$

or 
$$\int ae^{ia\tau} \psi_a da = \int e^{ia\tau} \alpha \psi_a da.$$



Since this equation holds for all values of  $\tau$ , we can equate coefficients of  $e^{ia\tau}$ , which gives

$$a\psi_a = \alpha\psi_a.$$

Thus  $\psi_a$  is an eigen- $\psi$  of  $\alpha$  belonging to the eigenvalue  $a$ . If we now put  $\tau = 0$  in (23), we obtain

$$\psi_0 = \int \psi_a da,$$

which gives an expansion for the arbitrary  $\psi_0$  in terms of the eigen- $\psi$ 's  $\psi_a$  in the form (21). If  $\psi_\tau$  were expressible as a Fourier series instead of the Fourier integral (23), we should get an expansion of  $\psi_0$  as a sum of eigen- $\psi$ 's in the form (20).

The weak point in the above argument is the assumption of a Fourier expansion for  $\psi_\tau$ . One case of failure would arise if the length of the vector  $\psi_\tau$  increased to infinity as  $\tau \rightarrow \infty$ , but this possibility can be ruled out with the help of the Hermitian condition for  $\alpha$ , which condition we have not yet used. Let  $\phi_\tau$  be the conjugate imaginary of  $\psi_\tau$ , satisfying the conjugate imaginary equation to (22), which is

$$\frac{d}{d\tau}\phi_\tau = -i\phi_\tau\alpha,$$

according to the theorem deduced in § 8 in connexion with equation (11). Then

$$\begin{aligned} \frac{d}{d\tau}(\phi_\tau\psi_\tau) &= \phi_\tau\frac{d\psi_\tau}{d\tau} + \frac{d\phi_\tau}{d\tau}\psi_\tau \\ &= \phi_\tau \cdot i\alpha\psi_\tau - i\phi_\tau\alpha \cdot \psi_\tau = 0. \end{aligned} \quad (24)$$

Thus the vector  $\psi_\tau$  remains of constant length and cannot increase to infinity.

The argument is not rigorous even yet, since the vector  $\psi_\tau$  might behave in other odd ways which would make its Fourier expansion impossible. In fact the expansion theorem is not true for every Hermitian operator, although it is true for most of the Hermitian operators met with in quantum mechanics. This results in lack of rigour in the theory from the mathematical standpoint, since we shall continually be requiring to use the expansion theorem in cases where it has not been proved. The situation is not so bad though, because there are usually physical grounds for telling when an application of the expansion theorem is permissible. The consistent development and physical interpretation of quantum mechanics require us to make the assumption that *only those Hermitian operators that satisfy the expansion theorem represent observables* and thus that *an arbitrary*



state is dependent on the eigenstates of any observable. Hence, if we know that a certain Hermitian operator represents some dynamical quantity which can be observed (for example if it represents the energy of some system) we may use the expansion theorem for this operator without fear of getting into error. Those linear operators appearing in the theory which do not represent observables will still represent functions of the dynamical variables, though these functions will not be directly observable.

## 11. Functions of an Observable

With the help of the algebraic operations of addition and multiplication we can give a meaning to those functions of linear operators that are expressible as power series and thus to the corresponding functions of observables. We can, however, get a much more general definition of a function of an observable by following out a suggestion mentioned in § 9.

Suppose we have any observable, represented by the Hermitian operator  $\alpha$ , and let  $\psi_a$  be one of its eigen- $\psi$ 's, belonging to the eigenvalue  $a$ , so that

$$\alpha\psi_a = a\psi_a. \quad (25)$$

A measurement of the observable when the system is in the state represented by  $\psi_a$  must certainly lead to the result  $a$ . We now require that a measurement of a function  $f$  of the observable, when the system is in this same state, shall certainly lead to the result  $f(a)$ ,  $f$  being any function such that  $f(a)$  has a meaning and is real. Thus we should expect the function of the observable to be represented by some linear operator  $f(\alpha)$  that satisfies

$$f(\alpha)\psi_a = f(a)\psi_a. \quad (26)$$

We take the condition that (26) always holds when (25) holds as the mathematical definition of  $f(\alpha)$ . It is easily seen that this definition is self-consistent, when applied to a set of eigen- $\psi$ 's of  $\alpha$  which are not independent, since such eigen- $\psi$ 's must all belong to the same eigenvalue. Thus, if we take for example three such eigen- $\psi$ 's,  $\psi_{1a}$ ,  $\psi_{2a}$  and  $\psi_{3a}$ , connected by the relation

$$\psi_{1a} = \psi_{2a} + \psi_{3a},$$

the definition would give us the same result if we operate on  $\psi_{1a}$  with  $f(\alpha)$  as if we operate on  $\psi_{2a}$  and  $\psi_{3a}$  and add. Also, the definition completely fixes the linear operator  $f(\alpha)$ , since it allows us to obtain the result of  $f(\alpha)$  operating on an arbitrary  $\psi$ -vector. We have only to



expand the arbitrary  $\psi$ -vector in terms of eigen- $\psi$ 's of  $\alpha$ , which expansion must be possible since  $\alpha$  represents an observable, and then to operate with  $f(\alpha)$  on each term separately in the expansion.†

We must now verify that the above-defined linear operator  $f(\alpha)$  can represent an observable. Evidently  $f(\alpha)$  satisfies the expansion theorem, since, as we see from (25) and (26), every eigen- $\psi$  of  $\alpha$  is an eigen- $\psi$  of  $f(\alpha)$ . It therefore only remains for us to verify the Hermitian condition. We can do this most conveniently by taking the form (10) for the Hermitian condition. Expanding the arbitrary  $\phi_x$  and  $\psi_y$  in terms of eigen- $\phi$ 's and eigen- $\psi$ 's of  $\alpha$  respectively,‡ we get equations of the form

$$\phi_x = \sum_a \phi_{xa}, \quad \psi_y = \sum_{a'} \psi_{ya'},$$

where  $\phi_{xa}$  is an eigen- $\phi$  belonging to the eigenvalue  $a$  and  $\psi_{ya'}$  is an eigen- $\psi$  belonging to the eigenvalue  $a'$ . Hence, with the help of (26),

$$\begin{aligned} \phi_x f(\alpha) \psi_y &= \sum_a \phi_{xa} f(\alpha) \sum_{a'} \psi_{ya'} = \sum_a \phi_{xa} \sum_{a'} f(a') \psi_{ya'} \\ &= \sum_{aa'} f(a') \phi_{xa} \psi_{ya'} = \sum_a f(a) \phi_{xa} \psi_{ya}, \end{aligned} \quad (27)$$

the orthogonality theorem being used in the last step. Interchanging the suffixes  $x$  and  $y$ , we get

$$\phi_y f(\alpha) \psi_x = \sum_a f(a) \phi_{ya} \psi_{xa}.$$

Now  $f(a)$  is, by hypothesis, a real number and from (4)

$$\phi_{xa} \psi_{ya} = \overline{\phi_{ya} \psi_{xa}},$$

so that

$$\phi_x f(\alpha) \psi_y = \overline{\phi_y f(\alpha) \psi_x}.$$

Hence  $f(\alpha)$  satisfies the condition that  $\alpha$  satisfies in (10) and must therefore be Hermitian.

We can now assume that the linear operator  $f(\alpha)$  represents that observable which is the function  $f$  of the observable represented by  $\alpha$ . In this way we can give a meaning to any real function  $f$  of an observable,

† We are here, for definiteness, taking the case when the expansion is in the form of a sum, as in (20), and not an integral, as in (21). We shall be continually doing this in the rest of this chapter and the next one. The change from sums to integrals involves only formal alterations in the theory, which alterations will be dealt with in Chapter IV.

‡ An expansion in terms of eigen- $\phi$ 's must, of course, be possible whenever the corresponding expansion for eigen- $\psi$ 's is possible. In these expansions we may assume, without loss of generality, that there is only one term corresponding to any eigenvalue, since if there were more than one, they could be lumped together to form a single term.



*provided only that the domain of existence of the function of a real variable  $f(x)$  includes all the eigenvalues of the observable.* If the domain of existence contains other points besides these eigenvalues, then the values of  $f(x)$  for these other points will not affect the function of the observable. The function need not be analytic or continuous. The eigenvalues of a function  $f$  of an observable are just the function  $f$  of the eigenvalues of the observable.

It is important to remember that the possibility of defining a function  $f$  of an observable requires the existence of a unique number  $f(x)$  for each value of  $x$  which is an eigenvalue of the observable. Thus the function must be single-valued and the function idea which we use corresponds to the one in the theory of functions of a real variable, rather than the one in the theory of functions of a complex variable. This may be illustrated by considering the question: When we have an observable  $f(A)$  which is a function of the observable  $A$ , is the observable  $A$  a function of the observable  $f(A)$ ? The answer to this is yes, if different eigenvalues  $a$  of  $A$  always correspond to different values of  $f(a)$ . If, however, there exist two different eigenvalues of  $A$ ,  $a_1$  and  $a_2$  say, such that  $f(a_1) = f(a_2)$ , then, corresponding to the eigenvalue  $f(a)$  of the observable  $f(A)$ , there will not be a unique eigenvalue of the observable  $A$  and the latter will not be a function of the observable  $f(A)$ .

It may easily be verified mathematically, from the definition, that the sum or product of two functions of an observable is a function of that observable and that a function of a function of an observable is a function of that observable. Also it is easily seen that the whole theory is symmetrical between  $\phi$ 's and  $\psi$ 's and that we could equally well work from the equations

$$\text{and} \quad \left. \begin{aligned} \phi_a \alpha &= a \phi_a \\ \phi_a f(\alpha) &= f(a) \phi_a \end{aligned} \right\} (28)$$

instead of from (25) and (26).

We shall conclude this section with a discussion of two examples which are of great practical importance, namely the reciprocal and the square root. The reciprocal of an observable exists when the observable does not have the eigenvalue zero. If the observable is represented by the Hermitian operator  $\alpha$ , the reciprocal observable will be represented by a Hermitian operator, which we call  $\alpha^{-1}$  or  $1/\alpha$ , satisfying

$$\alpha^{-1} \psi_a = a^{-1} \psi_a, \quad (29)$$



where  $\psi_a$  is an eigen- $\psi$  of  $\alpha$  belonging to the eigenvalue  $a$ . Hence

$$\alpha\alpha^{-1}\psi_a = \alpha a^{-1}\psi_a = \psi_a.$$

Since this holds for any eigen- $\psi$   $\psi_a$  and an arbitrary  $\psi$  can be expanded in terms of these eigen- $\psi$ 's, we must have

$$\alpha\alpha^{-1} = 1. \quad (30)$$

$$\text{Similarly,} \quad \alpha^{-1}\alpha = 1. \quad (31)$$

Either of these equations is sufficient to determine  $\alpha^{-1}$  completely, provided  $\alpha$  does not have the eigenvalue zero. To prove this in the case of (30), take the equation

$$\alpha\xi = 1$$

and multiply both sides on the left by the  $\alpha^{-1}$  defined by (29). The result is

$$\alpha^{-1}\alpha\xi = \alpha^{-1}$$

$$\text{and hence from (31)} \quad \xi = \alpha^{-1}.$$

Equations (30) and (31) can be used to define the reciprocal, when it exists, of a general linear operator  $\alpha$ , which need not be Hermitian and represent an observable. One of these equations by itself is then not necessarily sufficient.† If any two linear operators  $\alpha$  and  $\beta$  have reciprocals, their product  $\alpha\beta$  has the reciprocal

$$(\alpha\beta)^{-1} = \beta^{-1}\alpha^{-1}, \quad (32)$$

obtained by taking the reciprocal of each factor and reversing their order. We verify (32) by noting that its right-hand side gives unity when multiplied by  $\alpha\beta$ , either on the right or on the left. This reciprocal law for products can be immediately extended to more than two factors, i.e.,

$$(\alpha\beta\gamma\dots)^{-1} = \dots\gamma^{-1}\beta^{-1}\alpha^{-1}.$$

The square root of an observable exists when the observable has no negative eigenvalues. If the observable is represented by the Hermitian operator  $\alpha$ , the square root observable will be represented by a Hermitian operator, which we call  $\sqrt{\alpha}$  or  $\alpha^{\frac{1}{2}}$ , satisfying

$$\sqrt{\alpha}\psi_a = \pm\sqrt{a}\psi_a, \quad (33)$$

$\psi_a$  being an eigen- $\psi$  of  $\alpha$  belonging to the eigenvalue  $a$ . Hence

$$\sqrt{\alpha}\sqrt{\alpha}\psi_a = \sqrt{a}\sqrt{a}\psi_a = a\psi_a = \alpha\psi_a,$$

and since this holds for any eigen- $\psi$   $\psi_a$ , we must have

$$\sqrt{\alpha}\sqrt{\alpha} = \alpha. \quad (34)$$

† See, for example, the  $e^{i\omega}$  and  $e^{-i\omega}$  of § 36, equation (58).



On account of the ambiguity of sign in (33) there will be several square-root observables. To fix one of them we must specify a particular sign in (33) for each eigenvalue. This sign may vary irregularly from one eigenvalue to the next and equation (33) will always define a Hermitian operator  $\sqrt{\alpha}$  satisfying (34) and representing an observable which can legitimately be called a square root of our original observable. If there is an eigenvalue of  $\alpha$  with two or more independent eigen- $\psi$ 's belonging to it, then we must, according to our definition of a function, have the same sign in (33) for each of these eigen- $\psi$ 's. If we had different signs, however, equation (34) would still hold, and hence equation (34) by itself is not sufficient to define  $\sqrt{\alpha}$ , except in the special case when there is only one independent eigen- $\psi$  of  $\alpha$  belonging to any eigenvalue.

The number of different square roots of any observable which has no negative eigenvalues is  $2^n$ , where  $n$  is the total number of eigenvalues (or  $2^n - 1$  if one of the eigenvalues is zero). The square root mostly used in practice is the one for which the positive sign is always taken in (33). This one will be called the *positive square root*.

## 12. The General Physical Interpretation

The assumption that we introduced in § 9 to get a physical interpretation of our mathematics is of a rather special kind, since it can be used only in connexion with an equation of the special type (12). We need some more general assumption which will enable us to extract physical information from our mathematics even when we have no equation of the type (12).

In classical mechanics an observable always, as we say, 'has a value' for any particular state of the system. What is there in quantum mechanics corresponding to this? If we take any observable, represented by the Hermitian operator  $\alpha$  say, and any two states, represented by the vectors  $\phi_x$  and  $\psi_y$  say, then we can form the number  $\phi_x \alpha \psi_y$ . This number is not very closely analogous to the value which an observable can 'have' in the classical theory, for three reasons, namely, (i) it refers to *two* states of the system, while the classical value always refers to *one*, (ii) it is in general not a real number, and (iii) it is not uniquely determined by the observable and the states, since the vectors  $\phi_x$  and  $\psi_y$  contain arbitrary numerical factors. Even if we impose on  $\phi_x$  and  $\psi_y$  the condition that they shall be normalized, there will still be an undetermined factor of modulus



unity in  $\phi_x \alpha \psi_y$ . These three reasons cease to apply, however, if we take the two states to be identical. The number that we then get, namely  $\phi_x \alpha \psi_x$ , is necessarily real, as may be seen from equation (10) with the suffix  $y$  replaced by  $x$ . Also it is uniquely determined, with the help of the conditions that  $\phi_x$  and  $\psi_x$  are conjugate imaginary vectors and both normalized, since if we multiply  $\phi_x$  by the numerical factor  $e^{ic}$ ,  $c$  being some real number, we must multiply  $\psi_x$  by  $e^{-ic}$  and  $\phi_x \alpha \psi_x$  will be unaltered.

One might thus be inclined to make the tentative assumption that the observable represented by  $\alpha$  'has the value'  $\phi_x \alpha \psi_x$  for the state represented by  $\phi_x$  or  $\psi_x$ , in a sense analogous to the classical sense. This would not be satisfactory, though, for the following reason. Let us take a second observable, represented by the Hermitian operator  $\beta$ , and thus by the above assumption having the value  $\phi_x \beta \psi_x$  for this same state. We should expect, from classical analogy, that, for the same state again, the sum of the two observables would have a value equal to the sum of the values of the two observables separately and the product of the two observables would have a value equal to the product of the values of the two observables separately. Actually, the tentative assumption would give for the sum of the two observables the value  $\phi_x(\alpha + \beta)\psi_x$ , which is, in fact, equal to the sum of  $\phi_x \alpha \psi_x$  and  $\phi_x \beta \psi_x$ , but for the product it would give the value  $\phi_x \alpha \beta \psi_x$  or  $\phi_x \beta \alpha \psi_x$ , neither of which is connected in any simple way with  $\phi_x \alpha \psi_x$  and  $\phi_x \beta \psi_x$ .

However, since things go wrong only with the product and not with the sum, it would be reasonable to call  $\phi_x \alpha \psi_x$  the *average* value of the observable represented by  $\alpha$  for the state represented by  $\phi_x$  or  $\psi_x$ . This is because the average of the sum of two quantities must equal the sum of their averages, but the average of their product need not equal the product of their averages. We therefore make the general assumption that *if the measurement of the observable represented by  $\alpha$ , for the system in the state represented by  $\psi_x$ , is made a large number of times, the average of all the results obtained will be  $\phi_x \alpha \psi_x$ , provided  $\phi_x$  and  $\psi_x$  are normalized.* This assumption provides a general method for physical interpretation of the mathematics. We shall see a little later that the assumption of § 9 is deducible from this one.

The expression that an observable 'has a particular value' for a particular state is permissible in quantum mechanics in the special case when a measurement of the observable is certain to lead to the



particular value, so that an equation of the type (12) holds. It may easily be verified from the algebra that, with this restricted meaning for an observable 'having a value', if two observables have values for a particular state, then for this same state the sum of the two observables (if this sum is an observable<sup>†</sup>) has a value equal to the sum of the values of the two observables separately and the product of the two observables (if this product is an observable<sup>‡</sup>) has a value equal to the product of the values of the two observables separately.

In the general case we cannot speak of an observable having a value for a particular state, but we can speak of its having an average value for the state. We can go further and speak of the probability of its having any specified value for the state, meaning the probability of this specified value being obtained when one makes a measurement of the observable. This probability can be calculated from the general assumption for physical interpretation in the following way.

Take any observable, represented by the Hermitian operator  $\alpha$ , and any state, represented by the normalized  $\psi$ -vector  $\psi_x$ . Then the average value of the observable for the state will be  $\phi_x \alpha \psi_x$ . More generally, the average value of any function  $f$  of the observable will be  $\phi_x f(\alpha) \psi_x$ . This provides us with sufficient information to calculate the probability of the observable having any specified value. Suppose we expand  $\psi_x$  in terms of eigen- $\psi$ 's of  $\alpha$ , thus

$$\psi_x = \sum_a \psi_{xa}, \quad (35)$$

where  $\psi_{xa}$  is an eigen- $\psi$  of  $\alpha$  belonging to the eigenvalue  $a$ . Then, by the same analysis as led to (27) with the suffix  $y$  replaced by  $x$ , we obtain

$$\phi_x f(\alpha) \psi_x = \sum_a f(a) \phi_{xa} \psi_{xa}.$$

Now if  $P(a)$  is the probability of the result  $a$  being obtained from a measurement of the observable, the average value of the function  $f$  of the observable must be  $\sum_a f(a) P(a)$ , from the ordinary rules of probability, the summation being over all values  $a$  which are possible results of the measurement. Hence

$$\sum_a f(a) P(a) = \sum_a f(a) \phi_{xa} \psi_{xa}.$$

This equation holds for an arbitrary function  $f$ , so that  $f(a)$  can be an

<sup>†</sup> This is not obviously so, since the sum of the Hermitian operators representing the two observables may perhaps not satisfy the expansion theorem.

<sup>‡</sup> Here the Hermitian condition may fail, as well as the expansion theorem.



arbitrary number for each value of  $a$ . Hence we can equate coefficients of  $f(a)$ , which gives

$$P(a) = \phi_{xa} \psi_{xa}. \quad (36)$$

Thus the probability of the observable having any value  $a$  is equal to the square of the length of the corresponding eigen- $\psi$  in the expansion (35). If  $a$  is not an eigenvalue, there will be no eigen- $\psi$  corresponding to it in the expansion (35) and the probability must be zero. This proves the theorem, stated without proof in § 9, that every possible result of the measurement of an observable is one of its eigenvalues. It is easily confirmed that the expression (36) for  $P(a)$  gives unity for the total probability of the observable having as value any one of its eigenvalues. From the condition that  $\psi_x$  is normalized, we get

$$1 = \phi_x \psi_x = \sum_a \phi_{xa} \sum_{a'} \psi_{xa} = \sum_a \phi_{xa} \psi_{xa} = \sum_a P_a,$$

making use of the expansion (35) and its conjugate imaginary, and also of the orthogonality theorem.

We can now see that the assumption for physical interpretation made in § 9 is deducible from the one made in the present section. Let us apply the formula (36), which was obtained entirely from the physical interpretation of the present section, without the help of that of § 9, to the case of a state which is an eigenstate of the observable we are interested in. Then  $\psi_x$  will be an eigen- $\psi$  and the expansion on the right-hand side of (35) will contain only one term. This term will be normalized, so the square of its length will be unity. Formula (36) now tells us that the probability of the observable having any given value is unity if this value is the eigenvalue to which the state belongs and zero otherwise. This is just the converse of the initial assumption of § 9. The assumption itself can be deduced by a reversal of the argument.

We have been all the time taking the case when the expansion (35) is in the form of a sum and not an integral, and supposing, to agree with this, that the possible results of a measurement of the observable form a discrete set of numbers and not a continuous range. The case of integrals and continuous ranges will be dealt with in Chapter IV.

In practical applications of quantum mechanics it is nearly always more convenient to obtain the physical interpretation of the mathematics from formula (36) or something equivalent, instead of from a direct application of the expression for the average value of an observable.



### 13. Commutability and Compatibility

A state may be simultaneously an eigenstate of two observables. If the state is represented by the  $\psi$ -vector  $\psi$  and the observables are represented by the Hermitian operators  $\alpha$  and  $\beta$ , we should then have the equations

$$\alpha\psi = a\psi$$

$$\beta\psi = b\psi,$$

where  $a$  and  $b$  are numbers. We can now deduce

$$\alpha\beta\psi = \alpha b\psi = ab\psi = b\alpha\psi = \beta\alpha\psi,$$

or

$$(\alpha\beta - \beta\alpha)\psi = 0.$$

This suggests that the chances for the existence of a simultaneous eigenstate are most favourable if  $\alpha\beta - \beta\alpha = 0$  and the two observables commute. If they do not commute a simultaneous eigenstate is not impossible, but is rather exceptional. On the other hand, if they do commute there exist so many simultaneous eigenstates that, as will now be proved, *an arbitrary state is dependent on them*. We thus get a generalization of the expansion theorem of § 10.

Let  $\alpha$  and  $\beta$  be the Hermitian operators representing any two commuting observables. Take any eigen- $\psi$  of  $\alpha$ , say the eigen- $\psi$   $\psi_a$ , belonging to the eigenvalue  $a$ , and expand it in terms of eigen- $\psi$ 's of  $\beta$ , thus

$$\psi_a = \sum_b \psi_b, \quad (37)$$

where  $\psi_b$  is an eigen- $\psi$  of  $\beta$  belonging to the eigenvalue  $b$ . This expansion must be possible from § 10. From the equation

$$(\alpha - a)\psi_a = 0$$

we get

$$\sum_b (\alpha - a)\psi_b = 0. \quad (38)$$

Now  $\alpha\psi_b$  is an eigen- $\psi$  of  $\beta$  belonging to the eigenvalue  $b$ , since

$$\beta(\alpha\psi_b) = \alpha\beta\psi_b = \alpha b\psi_b = b(\alpha\psi_b).$$

Hence  $(\alpha - a)\psi_b$  is also an eigen- $\psi$  of  $\beta$  belonging to the eigenvalue  $b$ . Thus every term in the sum in (38) is an eigen- $\psi$  of  $\beta$  and each belongs to a different eigenvalue, since each term in the sum in (37) may be assumed to correspond to a different eigenvalue.† Now from a theorem of § 9, eigen- $\psi$ 's belonging to different eigenvalues are necessarily independent. It follows that every term in (38) vanishes separately. Thus

$$(\alpha - a)\psi_b = 0$$

† See second footnote on p. 39.



and each  $\psi_b$  is an eigen- $\psi$  of  $\alpha$  belonging to the eigenvalue  $a$  as well as being an eigen- $\psi$  of  $\beta$ . Equation (37) therefore gives  $\psi_a$  expanded in terms of simultaneous eigen- $\psi$ 's of  $\alpha$  and  $\beta$ . Since any  $\psi$  can be expanded in terms of  $\psi_a$ 's, it follows that any  $\psi$  can be expanded in terms of simultaneous eigen- $\psi$ 's of  $\alpha$  and  $\beta$ .

The converse theorem, which says that two observables must commute if an arbitrary  $\psi$  can be expanded in terms of their simultaneous eigen- $\psi$ 's, is also true. To prove it, let  $\alpha$  and  $\beta$  be the Hermitian operators representing the two observables and let  $\psi_{ab}$  be one of their simultaneous eigen- $\psi$ 's belonging to the eigenvalues  $a$  and  $b$ . We then have

$$(\alpha\beta - \beta\alpha)\psi_{ab} = (ab - ba)\psi_{ab} = 0.$$

Hence

$$(\alpha\beta - \beta\alpha)\psi = 0,$$

where  $\psi$  is any  $\psi$ -symbol that can be expanded in terms of the  $\psi_{ab}$ 's. If this is true for an arbitrary  $\psi$ , we can infer that

$$\alpha\beta - \beta\alpha = 0,$$

as required.

The idea of simultaneous eigen- $\psi$ 's may obviously be extended to more than two observables and the theorem proved above still holds, i.e., an arbitrary  $\psi$  can be expanded in terms of the simultaneous eigen- $\psi$ 's of any set of observables that commute, and also its converse. The same arguments used for the proof in the case of two observables are adequate for the general case; e.g., if we have three observables, represented by the Hermitian operators  $\alpha$ ,  $\beta$ ,  $\gamma$ , that commute, each with the other two, we can expand any simultaneous eigen- $\psi$  of  $\alpha$  and  $\beta$  in terms of eigen- $\psi$ 's of  $\gamma$  and then show that each of these eigen- $\psi$ 's of  $\gamma$  is also an eigen- $\psi$  of  $\alpha$  and  $\beta$ .

Two simultaneous eigen- $\psi$ 's must be orthogonal if the sets of eigenvalues to which they belong differ in any way.

Owing to the validity of the expansion theorem for two or more commuting observables, we can set up a theory of functions of two or more commuting observables, on the same lines as the theory of functions of a single observable given in § 11. If the commuting observables are represented by the Hermitian operators,  $\alpha$ ,  $\beta$ ,  $\gamma$ , ..., we define a general function  $f$  of them to be that observable represented by the Hermitian operator  $f(\alpha, \beta, \gamma, \dots)$  which satisfies

$$f(\alpha, \beta, \gamma, \dots)\psi_{abc\dots} = f(a, b, c, \dots)\psi_{abc\dots}, \quad (39)$$

where  $\psi_{abc\dots}$  is any simultaneous eigen- $\psi$  of  $\alpha$ ,  $\beta$ ,  $\gamma$ , ... belonging to the



eigenvalues  $a, b, c, \dots$ . Here  $f$  is any function such that  $f(a, b, c, \dots)$  is defined for all values of  $a, b, c, \dots$  which are eigenvalues of  $\alpha, \beta, \gamma, \dots$  respectively. The linear operator  $f(\alpha, \beta, \gamma, \dots)$  is completely defined by (39), since we can obtain the result of its operating on an arbitrary  $\psi$  by expanding this  $\psi$  in terms of the simultaneous eigen- $\psi$ 's  $\psi_{abc\dots}$  and operating on each term in the expansion separately.

We can now proceed to generalize the result (36). Suppose the normalized vector  $\psi_x$  representing any state to be expanded in terms of simultaneous eigen- $\psi$ 's of  $\alpha, \beta, \gamma, \dots$ , thus

$$\psi_x = \sum_{abc\dots} \psi_{xabc\dots} \quad (40)$$

Working from this equation instead of (35), we obtain, by an analogous argument to that which led to (36), that the probability for this state of the results  $a, b, c, \dots$  being obtained when measurements are made of the observables represented by  $\alpha, \beta, \gamma, \dots$  is

$$P(a, b, c, \dots) = \phi_{xabc\dots} \psi_{xabc\dots} \quad (41)$$

We can now conclude, in the first place, that *we can give a meaning to the probability of definite results being obtained for simultaneous measurements of several commuting observables*. This is not a trivial conclusion. In general one cannot make an observation on a system in a definite state without disturbing that state and spoiling it for the purposes of a second observation. One cannot then give any meaning at all to the two observations being made simultaneously. The above conclusion tells us, though, that in the special case when the two observables commute, the observations are to be considered as non-interfering or *compatible*, in such a way that one can give a meaning to the two observations being made simultaneously and can discuss the probability of any particular results being obtained. The two observations may, in fact, be considered as a single observation of a more complicated type, the result of which is expressible by two numbers instead of a single number. *From the point of view of general theory, any two or more commuting observables may be counted as a single observable, the result of a measurement of which consists of two or more numbers*. The states for which this measurement is certain to lead to one particular result are the simultaneous eigenstates.

The numerical value of the probability (41) is very important for applications of quantum mechanics.



### III

## REPRESENTATION THEORY FOR DISCRETE EIGENVALUES

### 14. The Bracket Notation

THE preceding chapter dealt with the fundamental laws governing states and observables in quantum mechanics and included all the axioms of the underlying mathematical formalism as well as the assumptions for physical interpretation of the mathematics. The present chapter and the following one will be concerned, not with making new laws and assumptions, but with systematizing and developing ideas and methods already introduced, and generally with arranging the theory in a form fitted for the subsequent applications. One matter that we must deal with is the setting up of a suitable notation for coordinates—a notation which can be consistently followed all through the future very extensive use of coordinates and is at the same time as simple and as easily remembered as possible.

In order to define a system of coordinates we must specify a set of  $\psi$ 's with the following properties. (i) They are all orthogonal to each other. (ii) Each of them is normalized. (iii) There are so many of them that an arbitrary  $\psi$  is dependent on them, so that if the space has a finite number of dimensions there must be the same number of these  $\psi$ 's. Such a set of  $\psi$ 's will be called a set of *basic*  $\psi$ 's for a coordinate system. The coordinates of any  $\psi$  will then be its coefficients when expanded in terms of the basic  $\psi$ 's. We shall denote a coordinate associated with a basic  $\psi$ ,  $\psi_r$  by the bracket expression  $(r|)$ . Thus

$$\psi = \sum_r \psi_r (r|).$$

We put the coordinates  $(r|)$  to the right of their corresponding  $\psi_r$ 's, in order to conform to a certain helpful style of writing, which will be developed as we go along. If we want to denote the coordinates of some particular  $\psi$ , specified by a suffix,  $x$  say, we put this suffix in the space to the right of the vertical line, so that the coordinate associated with the basic  $\psi$ ,  $\psi_r$  is  $(r|x)$ . Thus

$$\psi_x = \sum_r \psi_r (r|x). \quad (1)$$

The notation implies some kind of symmetry in the way a coordinate



$(r|x)$  depends on  $r$  and on  $x$ . We shall see in § 17 that there is such a symmetry.

The conjugate imaginaries of a set of basic  $\psi$ 's will be a set of basic  $\phi$ 's, defining a system of coordinates in the  $\phi$ -space, and the coordinates of an arbitrary  $\phi$  will be its coefficients when expanded in terms of the basic  $\phi$ 's. We denote the coordinate associated with the basic  $\phi, \phi_r$  by the bracket expression  $(|r)$ . Thus

$$\phi = \sum_r (|r)\phi_r,$$

our style of writing now requiring the coordinates to be put on the left. The corresponding coordinate of a particular  $\phi, \phi_x$  is denoted by  $(x|r)$ , so that

$$\phi_x = \sum_r (x|r)\phi_r. \quad (2)$$

The conditions of orthogonality and normalization which the basic  $\psi$ 's and  $\phi$ 's have to satisfy may all be expressed by the equation

$$\phi_r \psi_s = \delta_{rs}, \quad (3)$$

where the symbol  $\delta_{rs}$ , which we shall often use in the future, has the meaning

$$\left. \begin{aligned} \delta_{rs} &= 0 && \text{when } r \neq s \\ &= 1 && \text{when } r = s. \end{aligned} \right\} \quad (4)$$

If we multiply equation (1) by  $\phi_r$  on the left, we get, after first changing the dummy suffix  $r$  on the right-hand side into  $s$ ,

$$\begin{aligned} \phi_r \psi_x &= \sum_s \phi_r \psi_s (s|x) \\ &= (r|x), \end{aligned} \quad (5)$$

with the help of (3). This gives us an explicit expression for the coordinate  $(r|x)$  of  $\psi_x$ . A similar explicit expression for the coordinate  $(x|r)$  of  $\phi_x$  may be obtained by multiplying (2) by  $\psi_r$  on the right, the result being

$$\phi_x \psi_r = (x|r). \quad (6)$$

Other formulas which obviously hold are

$$(r|x) = \overline{(x|r)} \quad (7)$$

and

$$\phi_x \psi_y = \sum_r (x|r)(r|y). \quad (8)$$

The coordinates of a linear operator  $\alpha$  (Hermitian or not), which we previously denoted by  $\alpha_{rs}$ , will now be denoted by  $(r|\alpha|s)$ . These coordinates are defined by equations (6) and (5) of the preceding



chapter, which become, in our present notation

$$\left. \begin{aligned} \psi_b &= \alpha \psi_x \\ (r|b) &= \sum_s (r|\alpha|s)(s|x). \end{aligned} \right\} (9)$$

If the linear operator is considered to operate to the left on  $\phi$ -vectors, the corresponding equations are

$$\left. \begin{aligned} \phi_a &= \phi_y \alpha \\ (a|s) &= \sum_r (a|r)(r|\alpha|s). \end{aligned} \right\} (10)$$

By putting  $\psi_x$  in (9) equal to  $\psi_s$ , one of the basic  $\psi$ 's, with its  $s$ -th coordinate unity and all the others zero, we get

$$(r|b) = (r|\alpha|s).$$

Thus  $(r|\alpha|s)$  is the  $r$ -th coordinate of  $\alpha \psi_s$ , or, from (1),

$$\alpha \psi_s = \sum_r \psi_r (r|\alpha|s). \quad (11)$$

This equation may be taken as an alternative definition of the coordinates  $(r|\alpha|s)$ . The corresponding equation in terms of  $\phi$ 's, which may be derived by putting  $\phi_y = \phi_r$  in (10), is

$$\phi_r \alpha = \sum_s (r|\alpha|s) \phi_s. \quad (12)$$

By applying the formula (5) with  $\alpha \psi_s$  for  $\psi_x$ , we obtain an explicit expression for the  $r$ -th coordinate of  $\alpha \psi_s$ ,  $(r|\alpha|s)$ , namely the expression

$$(r|\alpha|s) = \phi_r \alpha \psi_s. \quad (13)$$

We could alternatively have obtained an explicit expression for  $(r|\alpha|s)$  by considering it to be a coordinate of  $\phi_r \alpha$ , as in (12), and using the formula (6). The result would have been the same. Our method of introducing a meaning for  $\alpha$  as an operator to the left on  $\phi$ -vectors was chosen so as to secure this agreement.

We have seen already that multiplication by a number,  $k$  say, is a special case of a linear operator. It may easily be deduced from (11) or (13) that the coordinates of this operator are

$$(r|k|s) = k \delta_{rs}. \quad (14)$$

All the coordinates vanish except those on the diagonal, i.e., those for which  $r = s$ , and the latter all have the value  $k$ . The identical operator, i.e., multiplication by unity, has for its coordinates just the numbers  $\delta_{rs}$ , forming a matrix which is called the unit matrix.

The general assumption of § 12 gives us a physical meaning for a diagonal coordinate  $(r|\alpha|r)$  of any linear operator  $\alpha$  that represents an



observable, namely  $(r|\alpha|r)$  is the average value of the observable represented by  $\alpha$  for the basic state represented by  $\psi_r$ .

A system of coordinates for  $\psi$ - and  $\phi$ -vectors and linear operators will be called in future a *representation*. The coordinates of any  $\psi$ - or  $\phi$ -vector or linear operator will be called the *representative* of that quantity and will be said to *represent* that quantity. They may also be called the representative of the corresponding state or observable.

## 15. Matrix Multiplication

Suppose we are given the representatives of two linear operators  $\alpha$  and  $\beta$ . What will be the representative of their product  $\alpha\beta$ ? Using the formula (11) twice, we get

$$\begin{aligned}\alpha\beta\psi_t &= \alpha \sum_s \psi_s(s|\beta|t) \\ &= \sum_{rs} \psi_r(r|\alpha|s)(s|\beta|t).\end{aligned}\tag{15}$$

But from this same formula we also have

$$\alpha\beta\psi_t = (\alpha\beta)\psi_t = \sum_r \psi_r(r|\alpha\beta|t).\tag{16}$$

Equating coefficients of  $\psi_r$  in (15) and (16), we obtain

$$(r|\alpha\beta|t) = \sum_s (r|\alpha|s)(s|\beta|t),\tag{17}$$

which gives us the representative of  $\alpha\beta$  in terms of those of  $\alpha$  and  $\beta$ .

If the representatives of our linear operators are regarded as forming matrices, equation (17) gives us the *matrix law of multiplication*, well known in pure mathematics. The element in the  $r$ -th row and  $t$ -th column of the product matrix is the sum of the product of each element in the  $r$ -th row of the first factor matrix with the corresponding element in the  $t$ -th column of the second factor matrix.

The second of equations (9) may also be regarded as an example of matrix multiplication. For this purpose the representative of any  $\psi$ -vector must be regarded as forming a matrix with just one column. The product of such a matrix with a square matrix, the square matrix being on the left, is again a matrix with a single column. Equations (9) now show that the single-column matrix representing  $\alpha\psi_x$  is equal to the product of the square matrix representing  $\alpha$  with the single-column matrix representing  $\psi_x$ . In a corresponding way the second of equations (10) becomes an example of matrix multiplication if we regard the representative of any  $\phi$ -vector as forming a matrix with a single row. Finally, equation (8) gives yet another example of



matrix multiplication, since its right-hand side may be regarded as a product of the single-row matrix representing  $\phi_x$  with the single-column matrix representing  $\psi_y$ , which product is, by the matrix law of multiplication, a matrix with one row and one column, i.e., an ordinary number.

The foregoing multiplication rules can be immediately extended to products of more than two factors, whether the factors are linear operators or  $\psi$ - or  $\phi$ -vectors. In every case the representative of the product is connected with those of the factors by the matrix law of multiplication. In consequence of this it is evident that the associative law of multiplication holds generally with all our symbols for linear operators and vectors, and that, for example,

$$\phi_x(\alpha\beta)\psi_y = (\phi_x\alpha)(\beta\psi_y).$$

In fact, all the laws of ordinary algebra hold with the exception of the commutative law of multiplication.

The rules of our style of writing have by now become fairly clear. When a summation is made over any variable, this variable occurs in two consecutive positions, on the extreme right of one factor and on the extreme left of the next following factor. The consistent use of this style makes it extremely easy to remember formulas such as (1), (2), (8), (9), (10), (17) and many others which will come later.

We define the conjugate complex  $\bar{\alpha}$  of any linear operator  $\alpha$  by

$$(r|\bar{\alpha}|s) = \overline{(s|\alpha|r)}, \quad (18)$$

i.e., the matrix representing  $\bar{\alpha}$  is obtained from that representing  $\alpha$  by the interchanging of rows and columns and the taking of the conjugate complex of each element. This rule, it should be noticed, is formally the same as that connecting the single-row matrix representing any  $\phi$  with the single-column matrix representing the conjugate imaginary  $\psi$ . We use the words 'conjugate complex' and not 'conjugate imaginary' when speaking of linear operators, because a linear operator and its conjugate complex are quantities of the same nature, which can be added together, and one can give a meaning to real and pure imaginary linear operators. A real linear operator, i.e., one equal to its conjugate complex, is, as we see from (18) with  $\bar{\alpha} = \alpha$ , just what we called in § 8 a Hermitian operator. The general linear operator corresponds to a complex function of the dynamical variables and a real or Hermitian linear operator corresponds to a real function



of the dynamical variables, which may be an observable. (It is an observable if it satisfies the expansion theorem.)

The conjugate complex of a product  $\alpha\beta$  of two linear operators may be obtained in the following way, from the formulas (18) and (17),

$$\begin{aligned}(r|\overline{\alpha\beta}|s) &= \overline{(s|\alpha\beta|r)} = \sum_t \overline{(s|\alpha|t)(t|\beta|r)} \\ &= \sum_t (r|\overline{\beta}|t)(t|\overline{\alpha}|s) = (r|\overline{\beta}\overline{\alpha}|s).\end{aligned}$$

Hence 
$$\overline{\alpha\beta} = \overline{\beta}\overline{\alpha}. \quad (19)$$

Thus to take the conjugate complex of a product of two linear operators, we must take the conjugate complex of each factor and reverse their order. The same rule applies to a product of three or more linear operators, as may be deduced by repeated applications of the rule for two linear operators, thus

$$\overline{\alpha\beta\gamma} = \overline{\alpha(\beta\gamma)} = \overline{\beta\gamma}\overline{\alpha} = \overline{\gamma}\overline{\beta}\overline{\alpha}.$$

The rule can easily be generalized still further to read—the conjugate complex or conjugate imaginary of any product of linear operators and  $\psi$ - and  $\phi$ -vectors is obtained by taking the conjugate complex or conjugate imaginary of each factor and reversing their order. The proof of the more general rule follows at once from the similarity of (7) and (18). This similarity allows us to infer, for instance by the same argument as led to (19), that the conjugate imaginary of  $\alpha\psi_x$  is  $\phi_x\overline{\alpha}$ . We have already had examples of the general rule in the preceding chapter in equations (4) and (10) and the result connected with equation (11), the last two examples being for the special case of  $\alpha$  real.

From (19) we see that if  $\alpha$  and  $\beta$  are two real linear operators, their product  $\alpha\beta$  need not be real. This product can be split up into a real part

$$\frac{1}{2}(\alpha\beta + \overline{\alpha\beta}) = \frac{1}{2}(\alpha\beta + \beta\alpha)$$

and a pure imaginary part

$$\frac{1}{2}(\alpha\beta - \overline{\alpha\beta}) = \frac{1}{2}(\alpha\beta - \beta\alpha).$$

Only when  $\alpha$  and  $\beta$  commute is the product  $\alpha\beta$  also real.

## 16. Eigen- $\psi$ 's as Basic $\psi$ 's

The connexion between an observable and the Hermitian operator that represents it in the sense of § 8 is so close that we can use the same letter to denote them both, without getting into confusion. We can, in fact, go further and count the observable and the Hermitian



operator as both the same thing, so that we say an observable *is* a Hermitian operator, which can operate either to the right on  $\psi$ -vectors or to the left on  $\phi$ -vectors. This provides a concise and convenient manner of speaking. A further rule of notation which we shall adopt is to denote an eigenvalue of an observable by the same letter that denotes the observable itself, with one or more primes attached. Thus the various eigenvalues of the observable  $\alpha$  will be denoted by  $\alpha', \alpha'', \dots, \alpha^{(n)}, \dots$ .

The representations that we have used up to the present have all been quite general. We must now consider the question of how to introduce a particular representation which shall be advantageous for some special problem. The idea for this is provided by the orthogonality theorem of § 9. Let us take some observable  $\xi$  and suppose for the present that its eigenvalues form a discrete set of numbers. Let us suppose further that it has only one independent eigen- $\psi$  belonging to any eigenvalue. If we now choose a normalized eigen- $\psi$  for each eigenvalue, we shall get a set of  $\psi$ 's, which are all orthogonal to each other and normalized and are such that an arbitrary  $\psi$  can be expanded in terms of them, so that they can be taken as the basic  $\psi$ 's of a representation.

There will be one basic  $\psi$  associated with each eigenvalue of  $\xi$ . The basic  $\psi$  associated with an eigenvalue  $\xi'$  we shall denote by  $\psi(\xi')$ . Also we shall use the eigenvalues associated with the various basic  $\psi$ 's as the labels for the corresponding coordinates, instead of the arbitrary labels  $r, s, t$  of the two preceding sections. Thus the coordinates of a  $\psi$  in our present representation will be written  $(\xi'|), (\xi''|), \dots$ , those of a  $\phi$ ,  $(|\xi'), (|\xi''), \dots$ , and those of a linear operator  $\alpha$  will be written like  $(\xi'|\alpha|\xi'')$ .

We can remove the restriction that there is only one independent eigen- $\psi$  of  $\xi$  belonging to any eigenvalue. If there are several independent eigen- $\psi$ 's belonging to some eigenvalue  $\xi'$ , we can choose out of all the eigen- $\psi$ 's belonging to this eigenvalue a set whose members are all normalized and orthogonal to each other and are such that any eigen- $\psi$  belonging to this eigenvalue can be expanded in terms of them. (This choice can in fact be made in an infinite number of ways.) Let us call the members of this set  $\psi(\xi'a'), \psi(\xi'a''), \dots$ . The whole assembly of  $\psi(\xi'a')$ 's for all  $\xi'$  and  $a'$  may now be taken as the basic  $\psi$ 's of a representation. The natural notation for coordinates in this representation is  $(\xi'a'|), (|\xi'a'), (\xi'a'|\alpha|\xi''a'')$ .



In this way we can set up a representation for which all the basic  $\psi$ 's are eigen- $\psi$ 's of some observable  $\xi$ . Let us see what the representative of  $\xi$  itself is in such a representation. We have from (11)

$$\xi \psi(\xi'' a'') = \sum_{\xi' a'} \psi(\xi' a') (\xi' a' | \xi | \xi'' a'').$$

But since  $\psi(\xi'' a'')$  is an eigen- $\psi$ ,

$$\xi \psi(\xi'' a'') = \xi'' \psi(\xi'' a'').$$

Equating the right-hand sides of these two equations, we obtain

$$(\xi' a' | \xi | \xi'' a'') = \xi'' \delta_{\xi' \xi''} \delta_{a' a''} = \xi' \delta_{\xi' \xi''} \delta_{a' a''}, \quad (20)$$

where the two-suffix  $\delta$ -symbols have the meaning (4).

The main feature of the matrix  $(\xi' a' | \xi | \xi'' a'')$  given by (20) is that all its matrix elements vanish except the diagonal ones, for which  $\xi' = \xi''$  and  $a' = a''$ . Such a matrix is called a *diagonal matrix*. Thus we have obtained a representation in which the observable  $\xi$  is represented by a diagonal matrix, or, as we may say for brevity, a representation in which  $\xi$  is diagonal. The elements on the diagonal are just the eigenvalues of  $\xi$ . In the next chapter we shall do the corresponding work for the case when the eigenvalues of  $\xi$  form a continuous range of numbers.† We shall then have the important and general result that *we can set up a representation in which any given observable is represented by a diagonal matrix, whose diagonal elements are just the eigenvalues of the observable*.

As an example of the usefulness of choosing a representation in which some given observable is diagonal, we shall prove the following theorem. *Any linear operator that commutes with an observable  $\xi$  commutes also with any function of  $\xi$* . The theorem is obviously true when the function is expressible as a power series. To prove it generally, let  $\omega$  be the linear operator, so that we have the equation

$$\xi \omega - \omega \xi = 0. \quad (21)$$

If we express this in terms of representatives in the above representation in which  $\xi$  is diagonal, we get from (17) and (20)

$$\sum_{\xi''' a'''} \{ \xi' \delta_{\xi' \xi'''} \delta_{a' a'''} (\xi''' a''' | \omega | \xi'' a'') - (\xi' a' | \omega | \xi''' a''') \xi''' \delta_{\xi' \xi'''} \delta_{a' a'''} \} = 0,$$

$$\text{or} \quad \xi' (\xi' a' | \omega | \xi'' a'') - (\xi' a' | \omega | \xi'' a'') \xi'' = 0. \quad (22)$$

The  $\xi$  in (21) is represented in (22) by the multiplying factor  $\xi'$  or  $\xi''$ .

† All the theorems and results of the present chapter will be obtained for the case of discrete eigenvalues only, the generalization to the case of continuous ranges of eigenvalues being left to the next chapter.



This illustrates a useful general rule which we can apply whenever we have to take the representative of an equation involving a diagonal observable. From (22) we now obtain

$$(\xi' a' | \omega | \xi'' a'') = 0 \quad \text{for} \quad \xi' \neq \xi'' \quad (23)$$

as the condition for  $\omega$  to commute with  $\xi$ . If  $f(\xi)$  denotes any function of  $\xi$ , its representative is, by the same argument as led to (20),

$$(\xi' a' | f(\xi) | \xi'' a'') = f(\xi') \delta_{\xi' \xi''} \delta_{a' a''}. \quad (24)$$

Using this, we obtain as the condition for  $\omega$  to commute with  $f(\xi)$ , by the same argument as led to (23),

$$(\xi' a' | \omega | \xi'' a'') = 0 \quad \text{for} \quad f(\xi') \neq f(\xi''). \quad (25)$$

Now (25) is obviously a consequence of (23) and so the theorem is proved.

As a special case of the theorem, we have the result that any observable that commutes with an observable  $\xi$  also commutes with any function of  $\xi$ . This result appears as a physical necessity when we identify, as in § 13, the condition of commutability of two observables with the condition of compatibility of the corresponding observations. Any observation that is compatible with the measurement of an observable  $\xi$  must also be compatible with the measurement of  $f(\xi)$ , since any measurement of  $\xi$  includes in itself a measurement of  $f(\xi)$ .

There is a converse theorem, which states that *if two observables  $\xi$  and  $g$  are such that any linear operator that commutes with  $\xi$  also commutes with  $g$ , then  $g$  is a function of  $\xi$* . To prove it, we take a general linear operator  $\omega$  that commutes with  $\xi$  and use again the above representation in which  $\xi$  is diagonal, so that we have equation (23). By hypothesis

$$g\omega - \omega g = 0,$$

and this, expressed in terms of representatives, gives us

$$\sum_{\xi''' a'''} \{ (\xi' a' | g | \xi''' a''') (\xi''' a''' | \omega | \xi'' a'') - (\xi' a' | \omega | \xi''' a''') (\xi''' a''' | g | \xi'' a'') \} = 0,$$

which reduces, with the help of (23), to

$$\sum_{a'''} \{ (\xi' a' | g | \xi'' a''') (\xi'' a''' | \omega | \xi'' a'') - (\xi' a' | \omega | \xi' a''') (\xi' a''' | g | \xi'' a'') \} = 0. \quad (26)$$

Now the numbers  $(\xi' a' | \omega | \xi' a'')$  are all arbitrary and independent, so that we can extract from (26) a great deal of information about the numbers  $(\xi' a' | g | \xi'' a'')$ . If we take  $\xi'$  differing from  $\xi''$  in (26), we see at once that

$$(\xi' a' | g | \xi'' a'') = 0 \quad \text{for} \quad \xi' \neq \xi''.$$



Further, putting  $\xi'' = \xi'$  in (26), we find that

$$(\xi'a'|g|\xi'a'') = 0 \quad \text{for } a' \neq a''$$

and

$$(\xi'a'|g|\xi'a') = (\xi'a''|g|\xi'a'').$$

Thus  $(\xi'a'|g|\xi''a'')$  is of the form

$$(\xi'a'|g|\xi''a'') = g(\xi')\delta_{\xi'\xi''}\delta_{a'a''}, \quad (27)$$

where  $g(\xi')$  is some function of  $\xi'$ , which has to be real, in order that the matrix representing the observable  $g$  may be Hermitian. Comparing (27) with (24), we see that the observable  $g$  is just that function of the observable  $\xi$  that  $g(\xi')$  is of the real variable  $\xi'$ .

A representation which we require to be such that a certain observable  $\xi$  is diagonal still has a great deal of arbitrariness left in it, if there are more than one independent eigen- $\psi$ 's of  $\xi$  belonging to any eigenvalue. We can reduce this arbitrariness by taking a second observable  $\eta$  that commutes with  $\xi$  and requiring the basic  $\psi$ 's to be simultaneous eigen- $\psi$ 's of  $\xi$  and  $\eta$ . We then get a representation in which both  $\xi$  and  $\eta$  are diagonal. If there are more than one independent simultaneous eigen- $\psi$ 's belonging to any pair of eigenvalues  $\xi', \eta'$ , we can introduce a third observable that commutes with both  $\xi$  and  $\eta$  and require the basic  $\psi$ 's to be simultaneous eigen- $\psi$ 's of all three, which will result in all three being diagonal. We can continue this process until eventually we have a representation for which the basic  $\psi$ 's are simultaneous eigen- $\psi$ 's of a set of commuting observables, the set including so many commuting observables that there is only one independent simultaneous eigen- $\psi$  of all of them belonging to any set of eigenvalues. Such a set of commuting observables will be called a *complete set of commuting observables*. This kind of representation is the most useful one in practice. In it each of the complete set of commuting observables will be diagonal. Further, the representation will be completely determined by the complete set of commuting observables, except for arbitrary phase factors arising from the fact that the basic  $\psi$ 's may be multiplied by arbitrary numbers of modulus unity without any of the conditions defining them being invalidated. We therefore conclude that *there exists a representation in which each of any set of commuting observables is simultaneously diagonal. If the set is a complete one, the representation is uniquely determined except for arbitrary phase factors in the basic  $\psi$ 's.*

Let the observables  $\xi_1, \xi_2, \dots, \xi_n$  form a complete commuting set and consider the representation in which they are diagonal. Each of the



basic  $\psi$ 's,  $\psi(\xi'_1 \xi'_2 \dots \xi'_n)$  say, will be specified by a set of eigenvalues  $\xi'_1, \xi'_2, \dots, \xi'_n$ . We can use these eigenvalues for labelling coordinates. A coordinate of a  $\psi$  or  $\phi$  will thus be written  $(\xi'_1 \xi'_2 \dots \xi'_n |)$  or  $(| \xi'_1 \xi'_2 \dots \xi'_n)$  respectively, which may be abridged to  $(\xi' |)$  or  $(| \xi')$  in work of a general theoretical nature. Similarly a coordinate of a linear operator  $\alpha$  will be written  $(\xi'_1 \xi'_2 \dots \xi'_n | \alpha | \xi''_1 \xi''_2 \dots \xi''_n)$ , or, abridged,  $(\xi' | \alpha | \xi'')$ . One of the  $\xi$ 's, say  $\xi_m$ , will itself be represented, according to (20), by

$$(\xi' | \xi_m | \xi'') = \xi'_m \delta_{\xi' \xi''}, \quad (28)$$

where  $\delta_{\xi' \xi''}$  has the meaning

$$\delta_{\xi' \xi''} = \delta_{\xi'_1 \xi''_1} \delta_{\xi'_2 \xi''_2} \dots \delta_{\xi'_n \xi''_n}. \quad (29)$$

The existence of arbitrary phase factors in the basic  $\psi$ 's means that we can multiply each  $\psi(\xi'_1 \xi'_2 \dots \xi'_n)$  by a numerical factor of the form  $e^{i\gamma'}$ , where  $\gamma' = \gamma(\xi'_1 \xi'_2 \dots \xi'_n)$  is any real function of the variables  $\xi'_1, \xi'_2, \dots, \xi'_n$ . Such a change in the representation would require us to multiply the representative  $(\xi' |)$  of any  $\psi$ -vector by  $e^{-i\gamma'}$ , the representative  $(| \xi')$  of any  $\phi$ -vector by  $e^{i\gamma'}$  and the representative  $(\xi' | \alpha | \xi'')$  of any linear operator  $\alpha$  by  $e^{i(\gamma'' - \gamma')}$ , where  $\gamma'' = \gamma(\xi''_1 \xi''_2 \dots \xi''_n)$ . A diagonal element of a linear operator remains unaltered, as is necessary on account of its physical meaning as an average, when the linear operator corresponds to an observable. For most purposes the arbitrary phase factors which exist in a representation are unimportant and trivial, so that we may count a representation as being completely determined by the observables that are diagonal in it. This fact is already implied in our notation, since the only indication in a representative of the representation to which it belongs are the letters denoting the observables that are diagonal.

If  $f(\xi'_1 \xi'_2 \dots \xi'_n) = f(\xi')$  denotes any function of the  $\xi$ 's, defined according to our general theory of functions of observables given in §§ 11, 13, then we find for its representative, by the same argument as led to (20) or (24),

$$(\xi' | f(\xi) | \xi'') = f(\xi') \delta_{\xi' \xi''}. \quad (30)$$

Thus any function of the observables  $\xi$  is represented by a diagonal matrix. Conversely, any Hermitian diagonal matrix represents a function of the  $\xi$ 's, since a general Hermitian diagonal matrix,  $g$  say, has for its elements

$$(\xi' | g | \xi'') = g(\xi') \delta_{\xi' \xi''},$$

where  $g(\xi')$  is some function of the variables  $\xi'$ , which has to be a real function from the Hermitian condition. This matrix must therefore represent that observable which is the function  $g$  of the observables  $\xi$ .



If  $\omega$  is a Hermitian operator that commutes with one of the  $\xi$ 's, say  $\xi_m$ , we see from (23) that  $(\xi'|\omega|\xi'')$  vanishes except when  $\xi'_m = \xi''_m$ . If  $\omega$  commutes with all the  $\xi$ 's,  $(\xi'|\omega|\xi'')$  must vanish except when  $\xi'_l = \xi''_l$  for all  $l$ . This means that  $\omega$  is a Hermitian diagonal matrix and hence represents a function of the  $\xi$ 's. We therefore have the theorem that *any Hermitian operator which commutes with each of a complete set of commuting observables is a function of those observables.*

It is easily seen that the theorem proved on pages 56 and 57, that any linear operator that commutes with an observable  $\xi$  commutes also with any function of  $\xi$ , and its converse are still valid when we replace the observable  $\xi$  by any set of commuting observables.

## 17. Transformation Theory

Let us take two representations, one with a complete set of commuting observables  $\xi_m$  diagonal and the other with another complete set of commuting observables  $\eta_l$  diagonal, and call them the  $\xi$ -representation and  $\eta$ -representation respectively. The basic  $\psi$ 's in the two representations we shall denote for brevity by  $\psi(\xi')$  and  $\psi(\eta')$ . An arbitrary  $\psi$  will now have the two representatives  $(\xi'|)$  and  $(\eta'|)$ , which are functions of the sets of variables  $\xi'_m$  and  $\eta'_l$  respectively, defined by

$$\psi = \sum_{\xi'} \psi(\xi')(\xi'|) \quad (31)$$

$$\text{and} \quad \psi = \sum_{\eta'} \psi(\eta')(\eta'|). \quad (32)$$

Since a  $\psi$  is completely determined by its representative in any one representation, there must be a connexion between the two representatives  $(\xi'|)$  and  $(\eta'|)$  such that either is determined by the other. We shall investigate the form of this connexion.

Each basic  $\psi$  of the  $\eta$ -representation,  $\psi(\eta')$ , will itself have a representative in the  $\xi$ -representation. We may write this representative  $(\xi'|\eta')$ , with  $\eta'$  on the right to show which  $\psi$  it represents. We shall then have

$$\psi(\eta') = \sum_{\xi'} \psi(\xi')(\xi'|\eta') \quad (33)$$

for the definition of  $(\xi'|\eta')$ . Substituting this expression for  $\psi(\eta')$  in the right-hand side of (32), we get

$$\psi = \sum_{\xi'\eta'} \psi(\xi')(\xi'|\eta')(\eta'|),$$

which gives, on comparison with (31)

$$(\xi'|) = \sum_{\eta'} (\xi'|\eta')(\eta'|). \quad (34)$$



This is the transformation equation which gives the  $\xi$ -representative of a  $\psi$  in terms of its  $\eta$ -representative. The corresponding equation which gives  $(\eta'|)$  in terms of  $(\xi'|)$  may be shown in the same way to be

$$(\eta'|) = \sum_{\xi'} (\eta'|\xi')(\xi'|), \quad (35)$$

where  $(\eta'|\xi')$  is the representative of the basic  $\psi$ ,  $\psi(\xi')$ , in the  $\eta$ -representation.

The two representatives  $(\xi'|)$  and  $(\eta'|)$  are thus linear functions of one another. The expressions  $(\xi'|\eta')$  and  $(\eta'|\xi')$  which enable us to pass from one to the other will be called transformation functions. They are each functions of the two sets of variables  $\xi'$  and  $\eta'$ . We can obtain an explicit expression for  $(\xi'|\eta')$  by multiplying equation (33) on the left by  $\phi(\xi')$ , corresponding to the way we obtained (5), the result being

$$(\xi'|\eta') = \phi(\xi')\psi(\eta'). \quad (36)$$

Similarly it may be shown that

$$(\eta'|\xi') = \phi(\eta')\psi(\xi'). \quad (37)$$

Hence  $(\xi'|\eta')$  and  $(\eta'|\xi')$  are conjugate complex numbers.

The transformation functions must satisfy certain conditions in order that (34) and (35) may be consistent. If we substitute for  $(\eta'|)$  in (34) its value given by (35), we get

$$(\xi'|) = \sum_{\eta', \xi''} (\xi'|\eta')(\eta'|\xi'')(\xi''|).$$

This must hold with  $(\xi'|)$  an arbitrary function of the variables  $\xi'$  and hence

$$\sum_{\eta'} (\xi'|\eta')(\eta'|\xi'') = \delta_{\xi'\xi''}, \quad (38)$$

the  $\delta$  symbol being defined by (29). The corresponding equation in which  $\xi$  and  $\eta$  have changed places, namely

$$\sum_{\xi'} (\eta'|\xi')(\xi'|\eta'') = \delta_{\eta'\eta''}, \quad (39)$$

may be deduced in the same way.

Let us now consider the transformation of the representatives of  $\phi$ 's. We may deal with these analogously to the  $\psi$ 's. We have as the equation which gives the representative  $(|\eta')$  of an arbitrary  $\phi$  in terms of its representative  $(|\xi')$

$$(|\eta') = \sum_{\xi'} (|\xi')(\xi'|\eta'),$$

where the quantity  $(\xi'|\eta')$  is now defined as the  $\eta$ -representative of



the basic  $\phi, \phi(\xi')$ , i.e. by the equation

$$\phi(\xi') = \sum_{\eta'} (\xi'|\eta')\phi(\eta').$$

If we multiply this equation by  $\psi(\eta'')$  on the right, we obtain, as an explicit expression for this  $(\xi'|\eta')$ ,

$$\phi(\xi')\psi(\eta'') = (\xi'|\eta''),$$

which is the same as (36). Hence this quantity  $(\xi'|\eta')$ , defined as the  $\eta$ -representative of  $\phi(\xi')$ , is the same as our previous one defined as the  $\xi$ -representative of  $\psi(\eta')$ , so that our notation of using the same symbol for them both is justified.

We can now understand the symmetry, referred to in § 14, in the way in which a coordinate  $(r|x)$  of any  $\psi, \psi_x$  involves on the one hand the parameter  $r$  specifying one particular coordinate and on the other the parameter  $x$  specifying the  $\psi$  whose coordinates we are considering. If  $\psi_x$  is normalized, we may consider it as one of the basic  $\psi$ 's of a new representation, and the coordinates  $(r|x)$  will then give us that part of the transformation function which refers to this one of the new basic  $\psi$ 's. Each  $(r|x)$  may be considered either as the  $r$ -th coordinate of  $\psi_x$  or as the  $x$ -th coordinate in the new representation of the basic  $\phi, \phi_r$  of the original representation.

The double meaning for a transformation function  $(\xi'|\eta')$  also enables us to understand better the significance of equations (38) and (39). If in (38) we regard  $(\xi'|\eta')$  as the  $\eta$ -representative of  $\phi(\xi')$  and  $(\eta'|\xi'')$  as the  $\eta$ -representative of  $\psi(\xi'')$ , the left-hand side of (38) becomes the  $\eta$ -representative of the product  $\phi(\xi')\psi(\xi'')$ , so the equation itself becomes equivalent to

$$\phi(\xi')\psi(\xi'') = \delta_{\xi'\xi''}. \quad (40)$$

Thus equation (38) just expresses, in terms of representatives in the  $\eta$ -representation, the condition (40), equivalent to (3), that the basic  $\psi$ 's of the  $\xi$ -representation are all orthogonal and normalized. Similarly equation (39) expresses, in terms of representatives in the  $\xi$ -representation, the condition that the basic  $\psi$ 's of the  $\eta$ -representation are all orthogonal and normalized. Equations (38) and (39), together with the condition that  $(\xi'|\eta')$  and  $(\eta'|\xi')$  must be conjugate complex, are the only conditions imposed on the transformation functions by general theoretical requirements.

Owing to the arbitrary phase factors occurring in representations, there will be a corresponding arbitrariness in the transformation



functions. If the basic  $\psi$ 's,  $\psi(\xi')$ ,  $\psi(\eta')$  are multiplied by  $\exp[if(\xi')]$ ,  $\exp[ig(\eta')]$  respectively,  $f$  and  $g$  being arbitrary real functions, the transformation function  $(\xi'|\eta')$  will get multiplied by

$$\exp\{-i[f(\xi')-g(\eta')]\}.$$

Thus the modulus of the transformation function is quite definite, the indeterminacy being only in its phase.

The connexion between the representatives of a linear operator  $\alpha$  in the two representations may easily be obtained in a variety of different ways. We can, for instance, use the explicit expression for the representative of  $\alpha$  given by equation (13). Applying this to the  $\xi$ -representation, we get

$$(\xi'|\alpha|\xi'') = \phi(\xi')\alpha\psi(\xi'').$$

If we now substitute for the right-hand side, which consists of the symbolic product of three factors, its representative in the  $\eta$ -representation, we get

$$(\xi'|\alpha|\xi'') = \sum_{\eta, \eta''} (\xi'|\eta')(\eta'|\alpha|\eta'')(\eta''|\xi''), \quad (41)$$

which gives the  $\xi$ -representative in terms of the  $\eta$ -representative. Similarly we may obtain the result

$$(\eta'|\alpha|\eta'') = \sum_{\xi, \xi''} (\eta'|\xi')(\xi'|\alpha|\xi'')(\xi''|\eta''), \quad (42)$$

giving the  $\eta$ -representative in terms of the  $\xi$ -representative. These are the transformation equations for the representatives of a linear operator. Either representative is a linear function of the other and the same transformation functions are required for passing from one to the other as for the representatives of  $\psi$ 's and  $\phi$ 's.

If we now take a third representation,  $\zeta$  say, we shall have transformation functions  $(\zeta'|\xi')$ ,  $(\xi'|\zeta')$  connecting it with the  $\xi$ -representation, and transformation functions  $(\zeta'|\eta')$ ,  $(\eta'|\zeta')$  connecting it with the  $\eta$ -representation. There are simple relations between the transformation functions. Equation (36), with  $\zeta$  instead of  $\eta$ , gives us

$$(\xi'|\zeta') = \phi(\xi')\psi(\zeta').$$

If we substitute for the right-hand side, which consists of the symbolic product of two factors, its representative in the  $\eta$ -representation, we get

$$(\xi'|\zeta') = \sum_{\eta} (\xi'|\eta')(\eta'|\zeta'). \quad (43)$$

The conjugate complex equation, which could be deduced indepen-



dently in the same way, is

$$(\zeta'|\xi') = \sum_{\eta'} (\zeta'|\eta')(\eta'|\xi'). \quad (44)$$

Equations (43) and (44) give the  $\xi, \zeta$  transformation functions in terms of the  $\xi, \eta$  and  $\eta, \zeta$  ones.

It is sometimes convenient to use a *mixed representative* of an observable or linear operator, that is to say, a representative in the form of a matrix whose rows and columns refer to two different representations. We define the mixed representative  $(\xi'|\alpha|\eta')$  of  $\alpha$  by

$$(\xi'|\alpha|\eta') = \sum_{\xi''} (\xi'|\alpha|\xi'')(\xi''|\eta'). \quad (45)$$

With the help of (41) we can express this mixed representative in terms of  $(\eta'|\alpha|\eta'')$ , the result being

$$(\xi'|\alpha|\eta') = \sum_{\eta''\eta'''\xi''} (\xi'|\eta'')(\eta''|\alpha|\eta''')(\eta'''|\xi'')(\xi''|\eta'),$$

which reduces, with the help of (39), to

$$\begin{aligned} (\xi'|\alpha|\eta') &= \sum_{\eta''\eta'''} (\xi'|\eta'')(\eta''|\alpha|\eta''')\delta_{\eta'''\eta'} \\ &= \sum_{\eta''} (\xi'|\eta'')(\eta''|\alpha|\eta'). \end{aligned} \quad (46)$$

Equation (46) may be taken as an alternative definition of the mixed representative  $(\xi'|\alpha|\eta')$ . By similar pieces of algebraic work one can verify that the ordinary representatives are given in terms of the mixed representative by

$$\begin{aligned} (\xi'|\alpha|\xi'') &= \sum_{\eta'} (\xi'|\alpha|\eta')(\eta'|\xi'') \\ (\eta'|\alpha|\eta'') &= \sum_{\xi'} (\eta'|\xi')(\xi'|\alpha|\eta''). \end{aligned}$$

The rows and the columns of a mixed representative are in general quite unrelated, so that no meaning can be given to its diagonal elements. The columns of one mixed representative may, however, refer to the same representation as, and thus be labelled to correspond to, the rows of a second mixed representative and we then have a multiplication law of the form

$$\sum_{\eta'} (\xi'|\alpha|\eta')(\eta'|\beta|\zeta') = (\xi'|\alpha\beta|\zeta'), \quad (47)$$

as is easily verified.

The identical operator has for its mixed representative  $(\xi'|1|\eta')$

$$(\xi'|1|\eta') = (\xi'|\eta'),$$

as follows at once from (45) or (46) and (14) with  $k = 1$ . We thus get a new meaning for the transformation function as the mixed repre-



sentative of the identical operator. Further, we obtain from (45) and (28)

$$(\xi'|\xi_m|\eta') = \xi'_m(\xi'|\eta'), \quad (48)$$

or more generally, from (45) and (30)

$$(\xi'|f(\xi)|\eta') = f(\xi')(\xi'|\eta').$$

Similarly, using (46) instead of (45), we find

$$\begin{aligned} (\xi'|\eta_l|\eta') &= (\xi'|\eta')\eta'_l \\ (\xi'|g(\eta)|\eta') &= (\xi'|\eta')g(\eta'). \end{aligned} \quad (49)$$

Finally, we have, using a multiplication law of the type (47),

$$\begin{aligned} (\xi'|f(\xi)g(\eta)|\eta') &= \sum_{\xi''} (\xi'|f(\xi)|\xi'') (\xi''|g(\eta)|\eta') \\ &= f(\xi')(\xi'|\eta')g(\eta'). \end{aligned} \quad (50)$$

## 18. Probability Amplitudes

In § 12 we obtained the probability of an observable having any specified value for a given state and in § 13 we generalized this result and obtained the probability of a set of commuting observables simultaneously having specified values for a given state. Let us now apply this result to a complete set of commuting observables, say the set of  $\xi$ 's which we have been dealing with already. According to equation (40) of § 13, we must take a normalized  $\psi$  representing the given state and expand it in terms of simultaneous eigen- $\psi$ 's of all the  $\xi$ 's. Equation (31) can now be used as this expansion, provided the  $\psi$  on the left-hand side of (31) is normalized, since on the right-hand side of (31) there is just one term corresponding to any set of eigenvalues  $\xi'$ . The difference in form between equation (40) of § 13 and equation (31) of § 17 consists merely in the simultaneous eigen- $\psi$ 's in the latter equation being written as normalized  $\psi$ 's with numerical coefficients. If we now apply the result (41) of § 13, we see that the probability, for the state represented by the  $\psi$  of (31), of the  $\xi$ 's simultaneously all having specified values  $\xi'$ , is

$$P(\xi') = |(\xi'|)|^2. \quad (51)$$

*Thus the probability of a complete set of commuting observables having specified values for a given state is equal to the square of the modulus of the coordinate, corresponding to these specified values, of a normalized  $\psi$  representing the state, in a representation in which each of the complete set of commuting observables is diagonal.*

There is therefore a simple physical meaning for the modulus of the representative of a normalized  $\psi$  in a representation in which each



of a complete set of commuting observables is diagonal. The existence of this physical meaning is the main reason why such representations are important. One may call the representative a *probability amplitude*, meaning by this something one must take the square of the modulus of to get an ordinary probability. There is no correspondingly simple physical meaning for the argument of the representative—indeed one cannot expect any, owing to the indeterminacy of this argument associated with the arbitrary phase factors of the representation. When the  $\psi$  is not normalized,  $|(\xi'|)|^2$  will still be proportional to the probability of the  $\xi$ 's having the values  $\xi'$ , the proportionality holding over all values of the  $\xi$ 's.

The probabilities that one calculates in practical problems in quantum mechanics are nearly always obtained as the squares of the moduli of probability amplitudes. Even when one is interested only in the probability of an incomplete set of commuting observables having specified values, it is usually necessary first to make the set a complete one by the introduction of some extra commuting observables and to obtain the probability of the complete set having specified values (as the square of the modulus of a probability amplitude), and then to sum over all possible values of the extra observables. A more direct application of formula (41) of § 13 is usually not practicable.

Let us now apply the formula (51) to a state which is one of the basic states of the  $\eta$ -representation, say the state represented by  $\psi(\eta')$ . This state is characterized physically by the property that a simultaneous measurement of all the  $\eta$ 's for it is certain to lead to the set of results  $\eta'$ . From (51) we see that the probability of the  $\xi$ 's having the values  $\xi'$  for this state is just  $|(\xi'|\eta')|^2$ , or the square of the modulus of the transformation function. The transformation function is now itself the probability amplitude. Since  $|(\xi'|\eta')|^2 = |(\eta'|\xi')|^2$ , we have the theorem of reciprocity,—*the probability of the  $\xi$ 's having the values  $\xi'$  for the state for which the  $\eta$ 's certainly have the values  $\eta'$  is equal to the probability of the  $\eta$ 's having the values  $\eta'$  for the state for which the  $\xi$ 's certainly have the values  $\xi'$* . The probability amplitude for either of these probabilities is the transformation function  $(\xi'|\eta')$  or  $(\eta'|\xi')$ .

The appearance of transformation functions as probability amplitudes results in the calculation of transformation functions being of practical importance. The general method of calculating the transformation function connecting two complete sets of commuting



observables consists in first obtaining the representatives of one set in the representation in which the other set are diagonal. When we know these representatives,  $(\xi'|\eta_l|\xi'')$  say, we can write down the equations

$$\sum_{\xi''} (\xi'|\eta_l|\xi'')(\xi''|\eta') = (\xi'|\eta')\eta'_l, \quad (52)$$

which follow at once from (45) with  $\alpha = \eta_l$  and (49), and proceed to solve them as algebraic equations in the unknowns  $(\xi'|\eta')$  and possibly also the  $\eta'$ 's. They are just of the standard form of equations in the theory of eigenvalues, equivalent to (18) of Chapter II. The general solution  $(\xi'|\eta')$  of (52) will contain an arbitrary function of the  $\eta'$ 's as a factor, and we must choose this function so as to satisfy the normalizing condition (39).

### 19. Example

As a simple example of the foregoing methods, let us consider three observables  $\sigma_x, \sigma_y, \sigma_z$  which satisfy the following relations

$$\left. \begin{aligned} \sigma_y \sigma_z - \sigma_z \sigma_y &= 2i\sigma_x \\ \sigma_z \sigma_x - \sigma_x \sigma_z &= 2i\sigma_y \\ \sigma_x \sigma_y - \sigma_y \sigma_x &= 2i\sigma_z \end{aligned} \right\} \quad (53)$$

$$\text{and} \quad \sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1. \quad (54)$$

This example is of importance for a study of the spin of the electron, as will be seen in § 39.

We note in the first place that equations (53) are permissible since, from the work at the end of § 15, their left-hand sides are pure imaginary and can therefore each be equated to  $i$  times an observable. We can get these equations into a simpler form with the help of some straightforward non-commutative algebra. From the first of equations (53) we obtain

$$\begin{aligned} 2i(\sigma_x \sigma_y + \sigma_y \sigma_x) &= (2i\sigma_x)\sigma_y + \sigma_y(2i\sigma_x) \\ &= (\sigma_y \sigma_z - \sigma_z \sigma_y)\sigma_y + \sigma_y(\sigma_y \sigma_z - \sigma_z \sigma_y) \\ &= -\sigma_z \sigma_y^2 + \sigma_y^2 \sigma_z \\ &= 0, \end{aligned}$$

from (54). Hence

$$\sigma_x \sigma_y = -\sigma_y \sigma_x.$$

Two observables or linear operators like these which satisfy the commutative law of multiplication except for a minus sign will be said to *anticommute*. Thus  $\sigma_x$  anticommutes with  $\sigma_y$ . From symmetry each of the three observables  $\sigma_x, \sigma_y, \sigma_z$  must anticommute with any other.



Equations (53) may therefore be written

$$\left. \begin{aligned} \sigma_y \sigma_z &= i\sigma_x = -\sigma_z \sigma_y \\ \sigma_z \sigma_x &= i\sigma_y = -\sigma_x \sigma_z \\ \sigma_x \sigma_y &= i\sigma_z = -\sigma_y \sigma_x \end{aligned} \right\} \quad (55)$$

and also from (54)

$$\sigma_x \sigma_y \sigma_z = i. \quad (56)$$

The three  $\sigma$ 's may be considered as the components of a vector  $\sigma$  in three-dimensional space and the algebraic equations which they satisfy would then be invariant under a rotation of axes. To verify this, let the components of  $\sigma$  referred to a new set of mutually perpendicular axes be

$$\sigma_1 = l_1 \sigma_x + m_1 \sigma_y + n_1 \sigma_z$$

$$\sigma_2 = l_2 \sigma_x + m_2 \sigma_y + n_2 \sigma_z$$

$$\sigma_3 = l_3 \sigma_x + m_3 \sigma_y + n_3 \sigma_z,$$

the  $l$ 's,  $m$ 's and  $n$ 's being the direction cosines of the new axes relative to the old ones. We then have from (54) and (55)

$$\begin{aligned} \sigma_1^2 &= (l_1 \sigma_x + m_1 \sigma_y + n_1 \sigma_z)^2 \\ &= l_1^2 \sigma_x^2 + m_1^2 \sigma_y^2 + n_1^2 \sigma_z^2 + l_1 m_1 (\sigma_x \sigma_y + \sigma_y \sigma_x) + \\ &\quad + m_1 n_1 (\sigma_y \sigma_z + \sigma_z \sigma_y) + n_1 l_1 (\sigma_z \sigma_x + \sigma_x \sigma_z) \\ &= l_1^2 + m_1^2 + n_1^2 = 1. \end{aligned}$$

Again

$$\begin{aligned} \sigma_2 \sigma_3 &= (l_2 \sigma_x + m_2 \sigma_y + n_2 \sigma_z)(l_3 \sigma_x + m_3 \sigma_y + n_3 \sigma_z) \\ &= l_2 l_3 \sigma_x^2 + m_2 m_3 \sigma_y^2 + n_2 n_3 \sigma_z^2 + l_2 m_3 \sigma_x \sigma_y + m_2 l_3 \sigma_y \sigma_x + \\ &\quad + m_2 n_3 \sigma_y \sigma_z + n_2 m_3 \sigma_z \sigma_y + n_2 l_3 \sigma_z \sigma_x + l_2 n_3 \sigma_x \sigma_z \\ &= l_2 l_3 + m_2 m_3 + n_2 n_3 + i(l_2 m_3 - m_2 l_3) \sigma_z + \\ &\quad + i(m_2 n_3 - n_2 m_3) \sigma_x + i(n_2 l_3 - l_2 n_3) \sigma_y \\ &= i(l_1 \sigma_x + m_1 \sigma_y + n_1 \sigma_z) = i\sigma_1. \end{aligned}$$

Thus  $\sigma_1, \sigma_2, \sigma_3$  satisfy relations of the same form as (54) and (55).

The eigenvalues of  $\sigma_x^2$  must be the squares of the eigenvalues of  $\sigma_x$ . Now from (54),  $\sigma_x^2$  has only the one eigenvalue unity and hence the eigenvalues of  $\sigma_x$  can be only 1 and  $-1$ . Both these numbers must be eigenvalues of  $\sigma_x$ , otherwise  $\sigma_x$  would be equal to a number and could not anticommute with anything. Similarly,  $\sigma_y$  and  $\sigma_z$  each have as their eigenvalues 1 and  $-1$ .

Let us set up a matrix representation for our observables  $\sigma$  and let us take  $\sigma_z$  to be diagonal. If there are no other independent observables besides the  $\sigma$ 's in our dynamical system, then  $\sigma_z$  by itself forms



a complete set of commuting observables, since the form of equations (54) and (55) is such that we cannot construct out of  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_z$  any new observable that commutes with  $\sigma_z$ . The diagonal elements of the matrix representing  $\sigma_z$  being the eigenvalues 1 and  $-1$  of  $\sigma_z$ , the matrix itself will be

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Let  $\sigma_x$  be represented by

$$\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}.$$

This matrix must be Hermitian, so that  $a_1$  and  $a_4$  must be real and  $a_2$  and  $a_3$  conjugate complex numbers. The equation  $\sigma_z \sigma_x = -\sigma_x \sigma_z$  gives us

$$\begin{pmatrix} a_1 & a_2 \\ -a_3 & -a_4 \end{pmatrix} = -\begin{pmatrix} a_1 & -a_2 \\ a_3 & -a_4 \end{pmatrix},$$

so that  $a_1 = a_4 = 0$ . Hence  $\sigma_x$  is represented by a matrix of the form

$$\begin{pmatrix} 0 & a_2 \\ a_3 & 0 \end{pmatrix}.$$

The equation  $\sigma_x^2 = 1$  now shows that  $a_2 a_3 = 1$ . Thus  $a_2$  and  $a_3$ , being conjugate complex numbers, must be of the form  $e^{i\alpha}$  and  $e^{-i\alpha}$  respectively, where  $\alpha$  is a real number, so that  $\sigma_x$  is represented by a matrix of the form

$$\begin{pmatrix} 0 & e^{i\alpha} \\ e^{-i\alpha} & 0 \end{pmatrix}.$$

Similarly it may be shown that  $\sigma_y$  is also represented by a matrix of this form. By suitably choosing the phase factors in the representation, which is not completely determined by the condition that  $\sigma_z$  shall be diagonal, we can arrange that  $\sigma_x$  shall be represented by the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The representative of  $\sigma_y$  is then determined by the equation  $\sigma_y = i\sigma_x \sigma_z$ . We thus obtain finally the three matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (57)$$

to represent  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_z$  respectively, which matrices satisfy all the algebraic relations (53), (54), (55), (56). The component of the vector  $\sigma$  in an arbitrary direction specified by the direction cosines  $l, m, n$ ,



namely  $l\sigma_x + m\sigma_y + n\sigma_z$ , is represented by

$$\begin{pmatrix} n & l-im \\ l+im & -n \end{pmatrix}. \quad (58)$$

In our representation with  $\sigma_z$  diagonal, the representative of a  $\psi$  will be written  $(\sigma'_z|)$  and will consist of just two numbers, corresponding to the two values  $+1$  and  $-1$  for  $\sigma'_z$ . These two numbers may be considered as forming a function of a variable  $\sigma'_z$  whose domain consists of only the two points  $+1$  and  $-1$ . The state for which  $\sigma_z$  has the value unity will be represented by the  $\psi$  whose representative is the function,  $f_\alpha(\sigma'_z)$  say, consisting of the pair of numbers 1, 0 and that for which  $\sigma_z$  has the value  $-1$  will be represented by the  $\psi$  whose representative is the function,  $f_\beta(\sigma'_z)$  say, consisting of the pair 0, 1. Any function of the variable  $\sigma'_z$ , i.e. any pair of numbers, can be expressed as a linear combination of these two. Thus *any state can be obtained by superposition of the two states for which  $\sigma_z$  equals  $+1$  and  $-1$  respectively*. For example, the state for which the component of  $\sigma$  in the direction  $l, m, n$ , represented by (58), has the value  $+1$  is represented by the  $\psi$  whose representative is the pair of numbers  $a, b$  which satisfy

$$\begin{pmatrix} n & l-im \\ l+im & -n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

or

$$na + (l-im)b = a$$

$$(l+im)a - nb = b.$$

Thus

$$\frac{a}{b} = \frac{l-im}{1-n} = \frac{1+n}{l+im}.$$

This state can be regarded as a superposition of the two states for which  $\sigma_z$  equals  $+1$  and  $-1$ , the relative weights in the superposition process being as

$$|a|^2 : |b|^2 = |l-im|^2 : (1-n)^2 = 1+n : 1-n.$$



## IV

### REPRESENTATION THEORY FOR CONTINUOUS EIGENVALUES

#### 20. Introduction of the $\delta$ function

IN the preceding chapter we saw the convenience of using a representation in which the basic  $\psi$ 's are eigen- $\psi$ 's of some chosen observable, or simultaneous eigen- $\psi$ 's of some chosen set of commuting observables. We considered, however, only the case when the chosen observables have as eigenvalues discrete sets of numbers, all our equations being written down for this case. It is possible for an observable to have as eigenvalues all numbers in a certain range, and in that case it becomes necessary to make some modification in the theory. From general physical grounds and from the possibility of regarding a continuous range of numbers as a limiting form of a discrete set whose density is increased indefinitely, one would expect the theory to run on somewhat parallel lines in the two cases. It would be desirable to have this parallelism as accurate as possible, and our development of the transformation theory for continuous ranges of eigenvalues will be made with this object in view.

Let us take an observable  $\xi$  with a continuous range of eigenvalues and suppose for the present that it has only one independent eigen- $\psi$  belonging to any eigenvalue. Then, ignoring for the present the question of normalization, we can take its eigen- $\psi$ 's,  $\psi(\xi')$ , as basic  $\psi$ 's of a representation. The number of these basic  $\psi$ 's, equal to the number of axes of our system of coordinates, is an infinity of a high order, equal to the number of points on a line, but this is not in itself a source of difficulty. The fundamental equation defining the representative of a  $\psi$ , corresponding to equation (1) of the preceding chapter, must now read

$$\psi_x = \int \psi(\xi') d\xi' (\xi'|x), \quad (1)$$

with an integral instead of a sum, the range of integration being understood to be the range of eigenvalues of  $\xi$ . (To conform to a neat style of writing in such equations it is desirable to put a differential element such as  $d\xi'$  between the two factors that involve the corresponding parameter  $\xi'$ .) The representative of  $\psi_x$ , namely  $(\xi'|x)$ , is now a function of the continuous variable  $\xi'$ .

We meet already with a difficulty in that not every  $\psi$  can be



expanded in the form (1), in spite of the expansion theorem requiring every  $\psi$  to be expressible as a linear function of the  $\psi(\xi')$ 's. An example of a  $\psi$  that cannot be expanded in the form (1) is one of the basic  $\psi$ 's itself, say  $\psi(\xi'')$ . Another example is the differential coefficient  $d\psi(\xi'')/d\xi''$ , when  $\psi(\xi'')$  involves the parameter  $\xi''$  in a sufficiently continuous way for this differential coefficient to exist. But in order that our theory for continuous eigenvalues may at all resemble our previous theory for discrete eigenvalues, it is necessary that such an expansion should be possible for every  $\psi$ , at least formally. We get over the difficulty by allowing the representative  $(\xi'|)$  to involve infinities and singularities of a certain type, chosen in just such a way as to make the expansion (1) always formally possible. This means allowing that  $(\xi'|)$  need not be a function of its variable  $\xi'$  according to the usual mathematical sense, which would require it to have a definite value for each value of its variable lying in a certain range, but may be something more general, which we call an *improper function* of the variable  $\xi'$ . Such an improper function may be pictured as the limit of a sequence of ordinary functions, corresponding to the fact that a  $\psi$  which cannot be expressed in the form (1) with  $(\xi'|)$  an ordinary function of  $\xi'$  may be regarded as the limit of a sequence of  $\psi$ 's that can.

The chief improper function which we shall have to deal with is the  $\delta$  function, defined by

$$\left. \begin{aligned} \int_{-\infty}^{\infty} \delta(x) dx &= 1 \\ \delta(x) &= 0 \text{ for } x \neq 0. \end{aligned} \right\} \quad (2)$$

To get a picture of  $\delta(x)$ , take a function of the real variable  $x$  which vanishes everywhere except inside a small domain, of length  $\epsilon$  say, surrounding the origin  $x = 0$ , and which is so large inside this domain that its integral over this domain is unity. The exact shape of the function inside this domain does not matter, provided there are no unnecessarily wild variations (for example provided the function is always of order  $\epsilon^{-1}$ ). Then in the limit  $\epsilon \rightarrow 0$  this function will go over into the  $\delta$  function.

The most important property of the  $\delta$  function is exemplified by the following equation,

$$\int_{-\infty}^{\infty} f(x) \delta(x) dx = f(0), \quad (3)$$



where  $f(x)$  is any continuous function of  $x$ . We can easily see the validity of this equation from the above picture of  $\delta(x)$ . The left-hand side of (3) can depend only on the values of  $f(x)$  very close to the origin, so that we may replace  $f(x)$  by its value at the origin,  $f(0)$ , without serious error. Equation (3) then follows from the first of equations (2). By making a change of origin in (3), we can deduce the formula

$$\int_{-\infty}^{\infty} f(x) \delta(x-a) dx = f(a), \quad (4)$$

where  $a$  is any real number. Thus *the process of multiplying a function of  $x$  by  $\delta(x-a)$  and integrating over all  $x$  is equivalent to the process of substituting  $a$  for  $x$* . The range of integration, of course, need not necessarily be from  $-\infty$  to  $\infty$ , but may be over any domain surrounding the critical point at which the  $\delta$  function does not vanish. In future the limits of integration will usually be omitted in such equations, it being understood that the domain of integration is a suitable one.

Formula (4) must hold equally well whether  $f$  is a scalar, or a vector or tensor function of  $x$ . By an application of (4) with  $f$  a  $\psi$ -vector, we see that

$$\psi(\xi'') = \int \psi(\xi') d\xi' \delta(\xi' - \xi''), \quad (5)$$

provided  $\psi(\xi')$  depends continuously on the variable  $\xi'$ . In this way we can express the basic  $\psi$ ,  $\psi(\xi'')$  in the form (1). The representative of  $\psi(\xi'')$  is just  $\delta(\xi' - \xi'')$ .

In order to express  $d\psi(\xi'')/d\xi''$  in the form (1), we have to use the derivative of the  $\delta$  function, which is another improper function, more improper than the  $\delta$  function itself. This derivative  $\delta'$  has the important property that, for any differentiable function  $f(x)$ ,

$$\int f(x) \delta'(x-a) dx = -f'(a). \quad (6)$$

We can verify this property either by integrating the left-hand side by parts and then applying (4) with  $f'(x)$  instead of  $f(x)$ , or by differentiating both sides of (4) with respect to  $a$ . The agreement of these two methods of verification provides evidence of the self-consistency of our use of improper functions. Formula (6) shows that *the process of multiplying a function of  $x$  by  $\delta'(x-a)$  and integrating over all  $x$  is equivalent to the process of differentiating it with respect to  $x$  and*



$$\frac{d\psi(\xi'')}{d\xi''} = - \int \psi(\xi') d\xi' \delta'(\xi' - \xi''), \quad (7)$$

which expresses the  $\psi$ -vector  $d\psi(\xi'')/d\xi''$  in the form (1) and shows that its representative is  $-\delta(\xi' - \xi'')$ . The higher derivatives of  $\psi(\xi'')$  with respect to  $\xi''$  can also be expressed in the form (1) with the help of the higher derivatives of the  $\delta$  function.

The foregoing work shows how the expansion of a  $\psi$  as an integral in the form (1) can be made of universal validity by the introduction of suitable improper functions. In this way we can get a foundation for the theory of representations in the case of continuous eigenvalues, corresponding to the foundation provided by equation (1) of Chapter III for the discrete case. Our definition and use of improper functions is not rigorous according to the standards of pure mathematics. It should be noticed, though, that an improper function can be given a rigorous meaning whenever it is a factor in an integrand. Now in the development of the theory, in every case where we have an improper function it will be something which is to be used finally only in integrands. We could therefore rewrite the theory in a form in which the improper functions appear all through only in integrands and could then eliminate the improper functions altogether and make the theory rigorous. The use of improper functions is thus not really connected with any essential lack of rigour in the theory. It is, rather, a convenient notation, enabling us to express in a concise form certain fundamental formulas which we could, if necessary, rewrite in a rigorous form, but only in a cumbersome way in which the parallelism with the case of discrete eigenvalues would be obscured. We shall confine our use of improper functions to such elementary equations that it will be obvious that the lack of rigour associated with them will not lead to a wrong result.

## 21. Properties of the $\delta$ function

An alternative way of defining the  $\delta$  function is as the differential coefficient  $\epsilon'(x)$  of the function  $\epsilon(x)$  given by

$$\begin{aligned} \epsilon(x) &= 0 & (x < 0) \\ &= 1 & (x > 0). \end{aligned}$$

We may verify formally that this is equivalent to the previous definition by substituting  $\epsilon'(x)$  for  $\delta(x)$  in the left-hand side of (3) and in-



tegrating by parts. We find, for  $g_1$  and  $g_2$  two positive numbers,

$$\begin{aligned}\int_{-g_2}^{g_1} f(x)\epsilon'(x) dx &= [f(x)\epsilon(x)]_{-g_2}^{g_1} - \int_{-g_2}^{g_1} f'(x)\epsilon(x) dx \\ &= f(g_1) - \int_0^{g_1} f'(x) dx \\ &= f(0),\end{aligned}$$

in agreement with (3). The  $\delta$  function appears whenever one differentiates a discontinuous function.

There are a number of elementary equations which one can write down about  $\delta$  functions. These equations are essentially rules of manipulation for algebraic work involving  $\delta$  functions. The meaning of any of these equations is that its two sides give equivalent results as factors in an integrand.

Examples of such equations are

$$\left. \begin{aligned}\delta(-x) &= \delta(x) \\ \delta'(-x) &= -\delta'(x),\end{aligned} \right\} (8)$$

$$x\delta(x) = 0, \quad (9)$$

$$x\delta'(x) = -\delta(x), \quad (10)$$

$$\delta(ax) = a^{-1}\delta(x) \quad (a > 0), \quad (11)$$

$$\delta(x^2 - a^2) = \frac{1}{2}a^{-1}\{\delta(x-a) + \delta(x+a)\} \quad (a > 0), \quad (12)$$

$$\int \delta(a-x) dx \delta(x-b) = \delta(a-b), \quad (13)$$

$$f(x)\delta(x-a) = f(a)\delta(x-a). \quad (14)$$

Equations (8), which merely state that  $\delta(x)$  is an even function of its variable  $x$ , are trivial. To verify (9) take any continuous function of  $x$ ,  $f(x)$ . Then

$$\int f(x)x\delta(x) dx = 0,$$

from (3). Thus  $x\delta(x)$  as a factor in an integrand is equivalent to zero, which is just the meaning of (9). Again, to verify (10) take any differentiable function  $f(x)$ . Thus, from (6) with  $a = 0$ ,

$$\begin{aligned}\int f(x)x\delta'(x) dx &= -\left[\frac{d}{dx}\{f(x)x\}\right]_{x=0} = -f(0) \\ &= -\int f(x)\delta(x) dx\end{aligned}$$

from (3). Thus  $x\delta'(x)$  as a factor in an integrand is equivalent to  $-\delta(x)$ , which is just the meaning of (10). (11) and (12) may be



verified by similar elementary arguments. To verify (13) take any continuous function of  $a$ ,  $f(a)$ . Then

$$\begin{aligned}\int f(a) da \int \delta(a-x) dx \delta(x-b) &= \int \delta(x-b) dx \int f(a) da \delta(a-x) \\ &= \int \delta(x-b) dx f(x) = \int f(a) da \delta(a-b).\end{aligned}$$

Thus the two sides of (13) are equivalent as factors in an integrand with  $a$  as variable of integration. It may be shown in the same way that they are equivalent also as factors in an integrand with  $b$  as variable of integration, so that equation (13) is justified from either of these points of view. Equation (14) is also easily justified, with the help of (4), from two points of view.

Equation (13) would be given by an application of (4) with  $f(x) = \delta(x-b)$ . We have here an illustration of the fact that we may often use an improper function as though it were an ordinary continuous function, without getting a wrong result. Another such illustration is obtained if we notice that the differentiation of equation (9) by the ordinary rules leads to equation (10).

A further example of a useful improper equation is that, for real  $a$ ,

$$\int_{-\infty}^{\infty} e^{iax} dx = 2\pi \delta(a). \quad (15)$$

From the standpoint of rigorous mathematics, the left-hand side of this equation is not a definite quantity at all, even when  $a$  differs from zero, since the integral even then does not converge when the limits of integration tend to  $\infty$  and  $-\infty$ , but oscillates. If, however, we fix the limits of integration at  $g_1$  and  $-g_2$  say, where  $g_1$  and  $g_2$  are both very large, and consider the dependence of the left-hand side of (15) on  $a$ , we see that it oscillates very rapidly about the value zero, except when  $a$  is very small. These oscillations will produce no effect in an integrand, and thus we can see the validity of (15) for  $a \neq 0$  according to our present standpoint. To verify (15) for  $a$  in the neighbourhood of zero, take any continuous function of  $a$ ,  $f(a)$ .

Then

$$\begin{aligned}\int_{-\infty}^{\infty} f(a) da \int_{-g}^g e^{iax} dx &= \int_{-\infty}^{\infty} f(a) da 2a^{-1} \sin ag \\ &= 2\pi f(0),\end{aligned}$$

in the limit when  $g$  tends to infinity. A rather more complicated argument shows that we get the same result if instead of the limits



$g$  and  $-g$  we put  $g_1$  and  $-g_2$ , and then let  $g_1$  and  $g_2$  tend to infinity in different ways. We can now see the equivalence of the two sides of (15) as factors in an integrand.

As an illustration of work with the  $\delta$  function, we may consider the differentiation of  $\log x$ . The usual formula

$$\frac{d}{dx} \log x = \frac{1}{x} \quad (16)$$

requires examination for the neighbourhood of  $x = 0$ . In order to make the reciprocal function  $1/x$  well-defined in the neighbourhood of  $x = 0$  (well-defined as an improper function) we must impose on it an extra condition, such as that its integral from  $-\epsilon$  to  $\epsilon$  vanishes. With this extra condition, the integral of the right-hand side of (16) from  $-\epsilon$  to  $\epsilon$  vanishes, while that of the left-hand side of (16) equals  $\log -1$ , so that (16) is not a correct equation. To correct it, we must remember that, taking principal values,  $\log x$  has a pure imaginary term  $i\pi$  for negative values of  $x$ . As  $x$  passes through the value zero this pure imaginary term vanishes discontinuously. The differentiation of this pure imaginary term gives us the result  $-i\pi\delta(x)$ , so that (16) should read

$$\frac{d}{dx} \log x = \frac{1}{x} - i\pi\delta(x). \quad (17)$$

The particular combination of reciprocal function and  $\delta$  function appearing in (17) plays an important part in the quantum theory of collision processes (see § 53).

## 22. Representations with One Continuous Parameter

Let us go back to the representation given by equation (1). Our problem now is to find a suitable way of fixing the length of the basic  $\psi$ 's  $\psi(\xi')$  as the usual normalization rule does not work here. We want some formula to replace (3) of Chapter III. We can attack this problem by referring to the physical meaning of the modulus of the representative  $(\xi'|)$  of a normalized  $\psi$ . If the eigenvalues  $\xi'$  are discrete, the square of this modulus,  $|(\xi')|^2$ , gives us, as we saw in § 18, the probability of  $\xi$  having the value  $\xi'$  for the state represented by this normalized  $\psi$ . If the eigenvalues  $\xi'$  are continuous, the probability of  $\xi$  having exactly the value  $\xi'$  for any physically obtainable state will be zero. The interesting quantity now will be the probability of  $\xi$  having a value lying within a specified range, say the small range from  $\xi'$  to  $\xi' + d\xi'$ . It would be convenient if we could arrange



so that *this probability is just*  $|(\xi'|)|^2 d\xi'$ . We should then have a close parallelism in the physical meanings of  $|(\xi'|)|$  in the two cases.

We want to arrange for the average value of any function  $f(\xi)$  of  $\xi$  to be  $\int f(\xi')|(\xi'|x)|^2 d\xi'$  for the state represented by any normalized  $\psi_x$ . Now from the general assumption of § 12, this average must be  $\phi_x f(\xi)\psi_x$ . We must therefore try to arrange so that

$$\phi_x f(\xi)\psi_x = \int f(\xi')|(\xi'|x)|^2 d\xi'. \quad (18)$$

From (1) and its conjugate imaginary

$$\phi_x = \int (x|\xi') d\xi' \phi(\xi'), \quad (19)$$

we get

$$\begin{aligned} \phi_x f(\xi)\psi_x &= \int (x|\xi') d\xi' \phi(\xi') f(\xi) \int \psi(\xi'') d\xi'' (\xi''|x) \\ &= \int (x|\xi') d\xi' \phi(\xi') f(\xi') \int \psi(\xi'') d\xi'' (\xi''|x) \end{aligned} \quad (20)$$

with the help of (28) of Chapter II. Now we want (18) to hold for an arbitrary function  $f$  and hence we can equate coefficients of  $f(\xi')$  on the right-hand sides of (18) and (20). This gives

$$|(\xi'|x)|^2 = (x|\xi')\phi(\xi') \int \psi(\xi'') d\xi'' (\xi''|x)$$

or 
$$(\xi'|x) = \int \phi(\xi')\psi(\xi'') d\xi'' (\xi''|x).$$

In order that this may hold for an arbitrary function  $(\xi'|x)$  of  $\xi'$ , we must have, from (4),

$$\phi(\xi')\psi(\xi'') = \delta(\xi' - \xi''). \quad (21)$$

Equation (21) is the fundamental formula which the basic  $\psi$ 's have to satisfy in the continuous case, corresponding to formula (3) of Chapter III in the discrete case. If written in rigorous mathematical notation without the  $\delta$  function, it would read

$$\phi(\xi')\psi(\xi'') = 0 \quad (\xi' \neq \xi'') \quad (22)$$

$$\int \phi(\xi')\psi(\xi'') d\xi'' = 1. \quad (23)$$

Equation (22) expresses that different basic  $\psi$ 's are orthogonal, exactly as in the discrete case. Equation (23) replaces the normalizing condition for the discrete case and may be called the *normalizing condition for a  $\psi$  labelled by a parameter that takes on a continuous range of values*. It should be remembered, though, that this involves a rather different meaning of the word 'normalizing' from what we had previously. A  $\psi(\xi')$  normalized according to (23) is not of unit length, but has an infinitely great length, as may be seen from (21).



Thus a  $\psi(\xi')$  normalized according to (23) would not be correctly normalized for the general physical interpretation of § 12 to be applicable, i.e. the average value of an observable  $\alpha$  for the state represented by such a  $\psi(\xi')$  would not be  $\phi(\xi')\alpha\psi(\xi')$ . It would be infinitely smaller. Still, the ratio of the average values of two observables  $\alpha$  and  $\beta$  would be the ratio of  $\phi(\xi')\alpha\psi(\xi')$  to  $\phi(\xi')\beta\psi(\xi')$ , and such ratios would be all one would usually wish to calculate about these average values. The state represented by a basic  $\psi$ ,  $\psi(\xi')$  is not of a kind that can actually exist. If an observable such as  $\xi$  with a continuous range of eigenvalues is measured for any actual state, the result must be distributed over a finite range according to some definite probability law, which range may be made as small as we please but cannot be contracted to a single point. The state represented by  $\psi(\xi')$  may, however, be regarded as a limit of actual states and as such it is a very useful theoretical abstraction.

We can now proceed with the development of the theory on parallel lines to Chapter III. From equation (1) with the suffix  $x$  replaced by  $y$  and equation (19), we get

$$\begin{aligned}\phi_x \psi_y &= \int \int (x|\xi') d\xi' \phi(\xi') \psi(\xi'') d\xi'' (\xi''|y) \\ &= \int \int (x|\xi') d\xi' \delta(\xi' - \xi'') d\xi'' (\xi''|y) \\ &= \int (x|\xi') d\xi' (\xi'|y),\end{aligned}\tag{24}$$

with the help of (21) and (4). This is the formula for the product of a  $\phi$  and a  $\psi$  in terms of their representatives. It corresponds to (8) of Chapter III, differing from that formula only in that the sum is replaced by an integral.

We define the representative of a linear operator  $\alpha$  by

$$\alpha\psi(\xi'') = \int \psi(\xi') d\xi' (\xi'|\alpha|\xi''),\tag{25}$$

corresponding to (11) of Chapter III. (An alternative definition, corresponding to (9) of Chapter III, would also be possible.) The representative  $(\xi'|\alpha|\xi'')$  is now a function of two variables  $\xi'$  and  $\xi''$  which can vary continuously. It is convenient to call such a function a 'matrix', using this word in a generalized sense, in order that we may be able to use the same terminology for the discrete and continuous cases. One of these generalized matrices cannot, of course, be written out as a two-dimensional array like an ordinary matrix, since the number of its rows and columns is an infinity equal to the



80 REPRESENTATION THEORY FOR CONTINUOUS EIGENVALUES § 22  
 number of points on a line. The law of multiplication for these generalized matrices is found to be, by an analogous piece of work to that leading to (17) of Chapter III,

$$(\xi'|\alpha\beta|\xi'') = \int (\xi'|\alpha|\xi''') d\xi''' (\xi'''|\beta|\xi''). \quad (26)$$

It is the same law as in the discrete case, except that the sum is replaced by an integral. More generally, one can easily verify that the whole theory of multiplication of representatives given in § 15, can be taken over if one just replaces sums by integrals all through. Equation (24) is an example of this. Further, the explicit expressions for representatives given by (5), (6) and (13) of § 14 have their analogues in the present theory. For example, corresponding to (5) of § 14, we get by multiplying (1) on the left by  $\phi(\xi')$

$$\begin{aligned} \phi(\xi')\psi_x &= \int \phi(\xi')\psi(\xi'') d\xi'' (\xi''|x) \\ &= \int \delta(\xi' - \xi'') d\xi'' (\xi''|x) \\ &= (\xi'|x) \end{aligned}$$

from (21) and (4). Similarly, corresponding to (13) of § 14, we find

$$(\xi'|\alpha|\xi'') = \phi(\xi')\alpha\psi(\xi''). \quad (27)$$

An element on the diagonal,  $(\xi'|\alpha|\xi')$ , is no longer, however, just the average value of the observable  $\alpha$  (when the linear operator  $\alpha$  represents an observable), for the basic state represented by  $\psi(\xi')$ , since the  $\phi$ - and  $\psi$ -vectors on the right-hand side of (27) are not correctly normalized for this to be so.

From (27) and (21) we find that the operator of multiplication by a number  $k$  is represented by

$$(\xi'|k|\xi'') = k\delta(\xi' - \xi''), \quad (28)$$

corresponding to (14) of Chapter III. The identical operator is represented by  $\delta(\xi' - \xi'')$ . For this reason the *unit matrix* in our present scheme of generalized matrices is defined as the matrix whose  $(\xi', \xi'')$  element is  $\delta(\xi' - \xi'')$ . As defined in this way, the unit matrix leaves unchanged any matrix when multiplied into it, on either the right- or left-hand side, according to the law of multiplication (26).

We define a *diagonal matrix* in our present scheme as one whose  $(\xi', \xi'')$  element is equal to some function of  $\xi'$  [or of  $\xi''$ , which would be equivalent, from (14)] multiplied into  $\delta(\xi' - \xi'')$  and this function, i.e. the coefficient of  $\delta(\xi' - \xi'')$ , we call a *diagonal element*. The



reason for this definition is that it is the widest one which gives to diagonal matrices the property of always commuting with one another, which property is a most fundamental one for diagonal matrices in the discrete case and is specially important in our theory of representations. It would not be sufficient to define a diagonal matrix merely as one whose  $(\xi', \xi'')$  elements all vanish except when  $\xi'$  differs infinitely little from  $\xi''$ , as that would include a matrix such as  $\delta'(\xi' - \xi'')$ , which, as is easily verified by the methods of § 21, does not commute with the matrix  $f(\xi')\delta(\xi' - \xi'')$  unless  $f(\xi')$  is a constant. With the above definition the unit matrix and the matrix  $(\xi'|k|\xi'')$  given by (28) are diagonal matrices. Further, the representative of  $\xi$  is also a diagonal matrix, since, as easily follows from (25) or (27),

$$(\xi'|\xi|\xi'') = \xi'\delta(\xi' - \xi''). \quad (29)$$

Thus choosing a representation in which the basic  $\psi$ 's are eigen- $\psi$ 's of  $\xi$  is equivalent to choosing a representation in which  $\xi$  is diagonal, in the case of continuous eigenvalues  $\xi'$  just as well as in the case of discrete eigenvalues, and the diagonal elements of the representative of  $\xi$  are in both cases just its eigenvalues  $\xi'$ . We can now see how the whole representation theory of the preceding chapter may be taken over to the case of a continuous range of basic states. We simply have to *replace sums by integrals and the two-suffix  $\delta$  symbol  $\delta_{\xi'\xi''}$  by the  $\delta$  function  $\delta(\xi' - \xi'')$ , all the way through.*

The transformation theory of § 17 may be taken over in the same way. If  $\eta$  is a second observable with continuous eigenvalues and we assume it for the present to have only one independent eigen- $\psi$  belonging to each eigenvalue, we can introduce a second representation in which the basic  $\psi$ 's are eigen- $\psi$ 's of  $\eta$  and in which, therefore,  $\eta$  is diagonal. There will then be transformation functions,  $(\xi'|\eta')$  and its conjugate complex  $(\eta'|\xi')$ , which enable one to pass from a  $\xi$ -representative to an  $\eta$ -representative by formulas analogous to those of § 17, with sums replaced by integrals. The conditions which the transformation functions have to satisfy will be

$$\left. \begin{aligned} \int (\xi'|\eta') d\eta' (\eta'|\xi'') &= \delta(\xi' - \xi'') \\ \int (\eta'|\xi') d\xi' (\xi'|\eta'') &= \delta(\eta' - \eta''), \end{aligned} \right\} \quad (30)$$

instead of (38) and (39) of § 17.

It would be quite possible for one representation to have continuous



82 REPRESENTATION THEORY FOR CONTINUOUS EIGENVALUES § 22  
eigenvalues and the other discrete.† We should then have similar transformation equations, with sums and integrals each occurring in the appropriate places. Instead of (30) we should have, if  $\xi'$  were continuous and  $\eta'$  discrete,

$$\left. \begin{aligned} \sum_{\eta'} (\xi' | \eta') (\eta' | \xi'') &= \delta(\xi' - \xi'') \\ \int (\eta' | \xi') d\xi' (\xi' | \eta'') &= \delta_{\eta' \eta''} \end{aligned} \right\} (31)$$

The physical interpretation of these transformation functions as probability amplitudes is evident. In the case of  $\xi'$  continuous and  $\eta'$  discrete,  $|(\xi' | \eta')|^2 d\xi'$  will be the probability of  $\xi$  having a value within the small range from  $\xi'$  to  $\xi' + d\xi'$  for the state for which  $\eta$  certainly has the value  $\eta'$ . Also  $|(\xi' | \eta')|^2$  will be proportional to the probability of  $\eta$  having the value  $\eta'$  for the state for which  $\xi$  certainly has the value  $\xi'$  (the proportionality holding for all values of  $\eta'$ ), but will not be equal to this probability, since the  $\phi$ -vector  $\phi(\xi')$  which is represented by  $(\xi' | \eta')$  in the  $\eta$ -representation is not properly normalized for this. In the case of both  $\xi'$  and  $\eta'$  continuous,  $|(\xi' | \eta')|^2 d\xi'$  will be proportional to the probability of  $\xi$  having a value within the small range  $\xi'$  to  $\xi' + d\xi'$  for the state for which  $\eta$  certainly has the value  $\eta'$ , and  $|(\xi' | \eta')|^2 d\eta'$  will be proportional to the probability of  $\eta$  having a value within the small range  $\eta'$  to  $\eta' + d\eta'$  for the state for which  $\xi$  certainly has the value  $\xi'$ .

### 23. General Representations

The work of the preceding section can be readily extended to the case when the observable  $\xi$  does not have only one independent eigen- $\psi$  belonging to any eigenvalue and when, following the method of the latter part of § 16, we take our basic  $\psi$ 's to be simultaneous eigen- $\psi$ 's of  $\xi$  and of one or more other observables which commute with  $\xi$  and with each other and which together with  $\xi$  form a complete set. Let us call the observables of this complete set  $\xi_1, \xi_2, \dots, \xi_n$ , and let us suppose each of them has a continuous range of eigenvalues. We can now make a straightforward generalization of our earlier theory, replacing the one-dimensional space of our former single variable  $\xi'$  by the  $n$ -dimensional space of the variables  $\xi'_1, \xi'_2, \dots, \xi'_n$ .

† If the number of basic  $\psi$ 's is finite in one representation, it must, of course, be finite and equal in any other representation, but it may be infinite enumerable in one representation and infinite equal to the number of points on a line in another. An example of this will be given in § 36.



Let us begin by obtaining the representative of one of the basic  $\psi$ 's,  $\psi(\xi_1'' \xi_2'' \dots \xi_n'')$ , or  $\psi(\xi'')$  say, for brevity. We note that

$$\psi(\xi'') = \int \int \dots \int \psi(\xi') d\xi_1' d\xi_2' \dots d\xi_n' \delta(\xi_1' - \xi_1'') \delta(\xi_2' - \xi_2'') \dots \delta(\xi_n' - \xi_n''), \quad (32)$$

as may be verified by carrying out the integrations one by one with the help of (4). This corresponds to (5) in the one-dimensional case. If we introduce the notation

$$\delta(\xi' - \xi'') = \delta(\xi_1' - \xi_1'') \delta(\xi_2' - \xi_2'') \dots \delta(\xi_n' - \xi_n''), \quad (33)$$

analogous to (29) of Chapter III, we find that the representative of  $\psi(\xi'')$  is just  $\delta(\xi' - \xi'')$ , a result which is formally the same as in the one-dimensional case. Also we have, for each  $m$  from 1 to  $n$ ,

$$\begin{aligned} \frac{\partial \psi(\xi'')}{\partial \xi_m''} = & - \int \int \dots \int \psi(\xi') d\xi_1' d\xi_2' \dots d\xi_n' \delta(\xi_1' - \xi_1'') \delta(\xi_2' - \xi_2'') \dots \\ & \dots \delta(\xi_{m-1}' - \xi_{m-1}'') \delta'(\xi_m' - \xi_m'') \delta(\xi_{m+1}' - \xi_{m+1}'') \dots \delta(\xi_n' - \xi_n''), \end{aligned} \quad (34)$$

as may be verified by carrying out the integrations one by one with the help of (4) and (6). The integrand in (34) differs from the integrand in (32) only in having the factor  $\delta'(\xi_m' - \xi_m'')$  instead of  $\delta(\xi_m' - \xi_m'')$ . Equation (34) is the generalization of (7) and gives us the representative of  $\partial \psi(\xi'') / \partial \xi_m''$ .

We again fix the lengths of our basic  $\psi$ 's in such a way as to give a simple physical meaning to the modulus of the representative  $(\xi_1' \xi_2' \dots \xi_n')$ , or  $(\xi'|)$  for brevity, of a normalized  $\psi$ . We can arrange for the probability of a simultaneous observation of all the  $\xi$ 's, for the state represented by this  $\psi$ , yielding for each  $\xi_m$  a result in the small domain between  $\xi_m'$  and  $\xi_m' + d\xi_m'$  to be

$$|(\xi'|)|^2 d\xi_1' d\xi_2' \dots d\xi_n'. \quad (35)$$

The condition for this turns out to be, with the notation (33), formally the same as (21). In fact if we use, in conjunction with (33), the notation of letting  $d\xi'$  denote the product  $d\xi_1' d\xi_2' \dots d\xi_n'$  and letting a single integral sign denote integration over all the variables  $\xi'$ , the equations and results of the preceding section will all apply, without formal alteration, to our present case. Thus, for example, the matrix law of multiplication (26) will still apply and the representative of a numerical multiplier will still be given by (28). Equation (29), to be made definite, should be rewritten

$$(\xi' | \xi_m | \xi'') = \xi_m' \delta(\xi' - \xi''), \quad (36)$$

and then applies to each  $\xi_m$  of the set  $\xi_1, \xi_2, \dots, \xi_n$ . In conformity with



our general plan, we must now define a diagonal matrix as one whose  $(\xi', \xi'')$  element is equal to some function of the  $\xi''$ 's, or of the  $\xi'$ 's, multiplied into  $\delta(\xi' - \xi'')$ . This results in each  $\xi_m$  being represented by a diagonal matrix. It can now easily be verified that all the theorems of § 16 are valid also for the case of continuous eigenvalues.

A further generalization which we must make in our theory is to allow some of the  $\xi$ 's in the complete commuting set to have discrete eigenvalues and others continuous eigenvalues. The alterations which this requires in the formalism are obvious. For each variable independently we must use either a sum or an integral, and either the two-suffix  $\delta$  symbol or the  $\delta$  function, according to whether it is discrete or continuous. We can make a transformation to another representation, in which each of a new complete set of commuting observables  $\eta_i$  is diagonal, with the help of transformation functions satisfying conditions which are an appropriate generalization of (30) or (31). There is no need for the number of discrete  $\eta$ 's, or the number of continuous  $\eta$ 's, to equal the number of discrete or continuous  $\xi$ 's respectively. In fact the total number of  $\eta$ 's in the  $\eta$ -set may differ from the total number of  $\xi$ 's in the  $\xi$ -set. This non-conservation in the total number of observables is connected with the circumstance, which we found in § 13, that two or more commuting observables may be counted as a single observable. We may at any time split up an observable into two commuting observables, a measurement of both of which is equivalent to a measurement of the original observable. For example, if the observable  $\alpha$  does not have the eigenvalue zero, we may split it up into  $\alpha^2$  and  $|\alpha|/\alpha$ , a measurement of both of these being equivalent to a measurement of  $\alpha$ .

We must make yet one more generalization in order to include all cases which may occur, namely we must allow any  $\xi$  to have as eigenvalues a discrete set of numbers together with a continuous range. This would give a representation theory in which sums and integrals occur added together in the formulas. The necessary alterations to be made in the formulas are obvious. For example, if we take the case of just one  $\xi$  with discrete eigenvalues denoted by  $\xi', \xi'', \dots$  and continuous eigenvalues denoted by  $\xi^{(a)}, \xi^{(b)}, \dots$ , the formula defining the representative of a  $\psi$ -vector, corresponding to (1), will be

$$\psi = \sum_{\xi'} \psi(\xi') (\xi' |) + \int \psi(\xi^{(a)}) d\xi^{(a)} (\xi^{(a)} |). \quad (37)$$

The representative of a  $\psi$  will now consist of the discrete set of



numbers  $(\xi'|)$  and the continuous range  $(\xi^{(a)})$ . These numbers may together be considered as forming a function of a variable whose domain consists of a discrete set of points together with a continuous range. They have the physical interpretation, when  $\psi$  is normalized, that  $|(\xi'|)|^2$  is the probability of  $\xi$  having the value  $\xi'$  and  $|(\xi^{(a)})|^2 d\xi^{(a)}$  is the probability of  $\xi$  having a value in the small range  $\xi^{(a)}$  to  $\xi^{(a)} + d\xi^{(a)}$ . The conditions for the basic  $\psi$ 's, corresponding to (21), will be

$$\left. \begin{aligned} \phi(\xi')\psi(\xi'') &= \delta_{\xi'\xi''} & \phi(\xi')\psi(\xi^{(b)}) &= 0 \\ \phi(\xi^{(a)})\psi(\xi'') &= 0 & \phi(\xi^{(a)})\psi(\xi^{(b)}) &= \delta(\xi^{(a)} - \xi^{(b)}). \end{aligned} \right\} \quad (38)$$

The representative of a linear operator will have four kinds of elements, typified by  $(\xi'|\alpha|\xi'')$ ,  $(\xi'|\alpha|\xi^{(b)})$ ,  $(\xi^{(a)}|\alpha|\xi'')$ ,  $(\xi^{(a)}|\alpha|\xi^{(b)})$ , and the matrix law of multiplication, corresponding to (26), will be

$$\begin{aligned} (\xi'|\alpha\beta|\xi'') &= \sum_{\xi'''} (\xi'|\alpha|\xi''')(\xi'''|\beta|\xi'') + \int (\xi'|\alpha|\xi^{(c)}) d\xi^{(c)} (\xi^{(c)}|\beta|\xi'') \\ (\xi'|\alpha\beta|\xi^{(b)}) &= \sum_{\xi'''} (\xi'|\alpha|\xi''')(\xi'''|\beta|\xi^{(b)}) + \int (\xi'|\alpha|\xi^{(c)}) d\xi^{(c)} (\xi^{(c)}|\beta|\xi^{(b)}) \\ (\xi^{(a)}|\alpha\beta|\xi'') &= \sum_{\xi'''} (\xi^{(a)}|\alpha|\xi''')(\xi'''|\beta|\xi'') + \int (\xi^{(a)}|\alpha|\xi^{(c)}) d\xi^{(c)} (\xi^{(c)}|\beta|\xi'') \\ (\xi^{(a)}|\alpha\beta|\xi^{(b)}) &= \sum_{\xi'''} (\xi^{(a)}|\alpha|\xi''')(\xi'''|\beta|\xi^{(b)}) + \int (\xi^{(a)}|\alpha|\xi^{(c)}) d\xi^{(c)} (\xi^{(c)}|\beta|\xi^{(b)}). \end{aligned}$$

In the general case of many  $\xi$ 's, the representative of a  $\psi$  will be a function whose domain may consist of several separate regions with different numbers of dimensions, and it may even be convenient to label the points of the various separate regions according to different schemes, referring perhaps to different sets of  $\xi$ 's as diagonal matrices.

## 24. The Weight Function

The foregoing discussion is sufficiently general to include all kinds of representations which can occur, but there is a purely formal generalization which it is sometimes desirable to make. This consists in introducing what is called a *weight function* into the expression (35) for the probability of the  $\xi$ 's having values in certain specified small ranges for any given state. We replace (35) by

$$|(\xi'|)|^2 \rho(\xi') d\xi'_1 d\xi'_2 \dots d\xi'_n, \quad (39)$$

where the weight function  $\rho(\xi')$  is any function of the variables  $\xi'$  which is greater than zero for all points in the domain of these variables. The use of a weight function is of no value from the point of view of general theory, but it is advantageous for certain special



applications, for the purpose of increasing the symmetry of the equations or of giving a more direct physical interpretation to  $|(\xi')|^2$  as a probability. For example, if two of the  $\xi$ 's are the angle variables  $\theta$  and  $\phi$  which fix some direction in space, it would be convenient to take as weight function  $\sin \theta'$ , in order to have the element of solid angle  $\sin \theta' d\theta' d\phi'$  in (39), so that we could interpret  $|(\xi')|^2$  directly as a probability per unit solid angle. It would be permissible to use a weight function also in the case of discrete eigenvalues, but there do not seem to be any examples in which it is then any help.

The effect of the introduction of the weight function in the various formulas is easily investigated. The two probabilities (39) and (35) must, of course, be the same, so that, putting  $\rho(\xi')$  equal to  $\rho'$  for brevity, we may take the  $(\xi'|)$  in (39) to be  $\rho'^{-\frac{1}{2}}$  times the  $(\xi'|)$  in (35). Formula (24) must now be replaced by

$$\phi_x \psi_y = \int (x|\xi') \rho' d\xi' (\xi'|y), \quad (40)$$

the extra factor  $\rho'$  being inserted in the integrand to compensate for each of the factors  $(x|\xi')$  and  $(\xi'|y)$  having  $\rho'^{-\frac{1}{2}}$  times its value in (24). We generalize this result to the assumption that *the weight function  $\rho'$  is to be inserted as a factor to the differential  $d\xi'$  in every formula involving an integration over the  $\xi''$ 's*, for example in (25) and (26). With this general assumption it is easily seen that all the formulas are still valid provided the various quantities they involve are all changed according to the following rules:

- (i) The representative  $(\xi'|)$  or  $(|\xi')$  of any  $\psi$ - or  $\phi$ -vector is multiplied by  $\rho'^{-\frac{1}{2}}$ , as we had above.
- (ii) Each basic  $\psi$ ,  $\psi(\xi')$ , and basic  $\phi$ ,  $\phi(\xi')$ , is multiplied by  $\rho'^{-\frac{1}{2}}$ .
- (iii) The representative  $(\xi'|\alpha|\xi'')$  of any linear operator is multiplied by  $\rho'^{-\frac{1}{2}} \rho''^{-\frac{1}{2}}$ .

Thus formula (28), for instance, gets altered to

$$(\xi'|k|\xi'') = k(\rho' \rho'')^{-\frac{1}{2}} \delta(\xi' - \xi'') = k \rho'^{-1} \delta(\xi' - \xi''),$$

and the representative of the identical operator becomes  $\rho'^{-1} \delta(\xi' - \xi'')$ .

For a certain type of general theoretical investigation the use of a continuous range of eigenvalues in the representation is extremely inconvenient and it becomes desirable, and is permissible, to replace the continuous range by a discrete set of points lying very close to one another over the whole range and eventually to pass to the limit when the density of these points is everywhere infinite. This proce-



ture is equivalent to the introduction of a certain weight function, which tends to infinity in the limit. Let the number of discrete points in the small domain  $\xi'$  to  $\xi' + d\xi'$  (which may be either one-dimensional or many-dimensional) be  $s' d\xi'$ , where  $s' = s(\xi')$  is any function of  $\xi'$  which is everywhere large. Thus  $s'$  is the density of the discrete points. The general formula connecting a sum over the discrete points with an integral over the continuous range is now

$$\sum_{\xi'} F(\xi') = \int F(\xi') s' d\xi', \quad (41)$$

which shows that the discrete representation is equivalent to that continuous representation in which the weight function  $s$  has been introduced. We may now use the rules (i), (ii), (iii), with  $\rho$  replaced by  $s$ . We shall have, for example, using the suffix  $D$  to denote a representative in the discrete representation

$$\left. \begin{aligned} (\xi'|)_D &= s'^{-\frac{1}{2}}(\xi'|), & (|\xi')_D &= (|\xi') s'^{-\frac{1}{2}} \\ (\xi'|\alpha|\xi'')_D &= s'^{-\frac{1}{2}}(\xi'|\alpha|\xi'') s''^{-\frac{1}{2}}, \end{aligned} \right\} (42)$$

and further

$$\sum_{\xi'} (|\xi')_D (\xi'|)_D = \sum_{\xi'} (|\xi') s'^{-1} (\xi'|) = \int (|\xi') d\xi' (\xi'|) \quad (43)$$

from (41).



## THE QUANTUM CONDITIONS

## 25. Poisson Brackets

OUR work so far has consisted in setting up a general mathematical scheme connecting states and observables in quantum mechanics. One of the dominant features of this scheme is that observables, and dynamical variables in general, appear in it as quantities which do not obey the commutative law of multiplication. It now becomes necessary for us to obtain equations to replace the commutative law of multiplication, equations that will tell us the value of  $\xi\eta - \eta\xi$  when  $\xi$  and  $\eta$  are any two observables or dynamical variables. Only when such equations are known shall we have a complete scheme of mechanics with which to replace classical mechanics. These new equations are called *quantum conditions* or *commutability relations*.

The problem of finding quantum conditions is not of such a general character as those we have been concerned with up to the present. It is instead a special problem which presents itself with each particular dynamical system one is called upon to study. There is, however, a fairly general method of obtaining quantum conditions, applicable to a very large class of dynamical systems. This is the method of classical analogy and will form the main theme of the present chapter. Those dynamical systems to which this method is not applicable must be treated individually and special considerations used in each case.

The value of classical analogy in the development of quantum mechanics depends on the fact that classical mechanics provides a valid description of dynamical systems under certain conditions, when the particles and bodies composing the systems are sufficiently massive for the disturbance accompanying an observation to be negligible. Classical mechanics must therefore be a limiting case of quantum mechanics. We should thus expect to find that important concepts in classical mechanics correspond to important concepts in quantum mechanics, and, from an understanding of the general nature of the analogy between classical and quantum mechanics, we may hope to get laws and theorems in quantum mechanics appearing as simple generalizations of well-known results in classical mechanics; in particular we may hope to get the quantum conditions appearing



as a simple generalization of the classical law that all dynamical variables commute.

One of the fundamental ideas of classical mechanics is that of generalized coordinates  $q_r$  and their canonically conjugate momenta  $p_r$ . An idea possibly still more fundamental, however, is that of the Poisson Bracket. Any two dynamical variables  $\xi$  and  $\eta$  have a P.B. (Poisson Bracket) which we shall denote by  $[\xi, \eta]$ , defined by

$$[\xi, \eta] = \sum_r \left\{ \frac{\partial \xi}{\partial q_r} \frac{\partial \eta}{\partial p_r} - \frac{\partial \xi}{\partial p_r} \frac{\partial \eta}{\partial q_r} \right\}, \quad (1)$$

$\xi$  and  $\eta$  being regarded as functions of the canonical coordinates and momenta  $q_r$  and  $p_r$  for the purpose of the differentiations. The P.B. owes its importance to its being invariant under a contact transformation, i.e. a transformation to a new set of canonical coordinates and momenta, so that it depends only on the two dynamical variables  $\xi$  and  $\eta$  to which it refers and is independent of which set of canonical coordinates and momenta one is using. The main properties of P.B.'s, which follow at once from their definition (1), are

$$[\xi, \eta] = -[\eta, \xi] \quad (2)$$

$$[\xi, c] = 0 \quad (3)$$

where  $c$  is a number,

$$\left. \begin{aligned} [\xi_1 + \xi_2, \eta] &= [\xi_1, \eta] + [\xi_2, \eta] \\ [\xi, \eta_1 + \eta_2] &= [\xi, \eta_1] + [\xi, \eta_2] \end{aligned} \right\} \quad (4)$$

$$\left. \begin{aligned} [\xi_1 \xi_2, \eta] &= \sum_r \left\{ \left( \frac{\partial \xi_1}{\partial q_r} \xi_2 + \xi_1 \frac{\partial \xi_2}{\partial q_r} \right) \frac{\partial \eta}{\partial p_r} - \left( \frac{\partial \xi_1}{\partial p_r} \xi_2 + \xi_1 \frac{\partial \xi_2}{\partial p_r} \right) \frac{\partial \eta}{\partial q_r} \right\} \\ &= [\xi_1, \eta] \xi_2 + \xi_1 [\xi_2, \eta] \\ [\xi, \eta_1 \eta_2] &= [\xi, \eta_1] \eta_2 + \eta_1 [\xi, \eta_2]. \end{aligned} \right\} \quad (5)$$

Also the identity

$$[\xi, [\eta, \zeta]] + [\eta, [\zeta, \xi]] + [\zeta, [\xi, \eta]] = 0 \quad (6)$$

is easily verified. Equations (4) express that the P.B.  $[\xi, \eta]$  involves  $\xi$  and  $\eta$  linearly, while equations (5) correspond to the ordinary rules for differentiating a product.

Let us try to introduce a quantum P.B. which shall be the analogue of the classical one. We assume the quantum P.B. to satisfy all the conditions (2) to (6), it being now necessary that the order of the factors  $\xi_1$  and  $\xi_2$  in the first of equations (5) should be preserved throughout the equation, as in the way we have here written it, and



similarly for the  $\eta_1$  and  $\eta_2$  in the second of equations (5). These conditions are already sufficient to determine the form of the quantum P.B. uniquely, as may be seen from the following argument. We can evaluate the P.B.  $[\xi_1 \xi_2, \eta_1 \eta_2]$  in two different ways, since we can use either of the two formulas (5) first, thus,

$$\begin{aligned} [\xi_1 \xi_2, \eta_1 \eta_2] &= [\xi_1, \eta_1 \eta_2] \xi_2 + \xi_1 [\xi_2, \eta_1 \eta_2] \\ &= \{[\xi_1, \eta_1] \eta_2 + \eta_1 [\xi_1, \eta_2]\} \xi_2 + \xi_1 \{[\xi_2, \eta_1] \eta_2 + \eta_1 [\xi_2, \eta_2]\} \\ &= [\xi_1, \eta_1] \eta_2 \xi_2 + \eta_1 [\xi_1, \eta_2] \xi_2 + \xi_1 [\xi_2, \eta_1] \eta_2 + \xi_1 \eta_1 [\xi_2, \eta_2] \end{aligned}$$

and

$$\begin{aligned} [\xi_1 \xi_2, \eta_1 \eta_2] &= [\xi_1 \xi_2, \eta_1] \eta_2 + \eta_1 [\xi_1 \xi_2, \eta_2] \\ &= [\xi_1, \eta_1] \xi_2 \eta_2 + \xi_1 [\xi_2, \eta_1] \eta_2 + \eta_1 [\xi_1, \eta_2] \xi_2 + \eta_1 \xi_1 [\xi_2, \eta_2]. \end{aligned}$$

Equating these two results, we obtain

$$[\xi_1, \eta_1] (\xi_2 \eta_2 - \eta_2 \xi_2) = (\xi_1 \eta_1 - \eta_1 \xi_1) [\xi_2, \eta_2].$$

Since this condition holds with  $\xi_1$  and  $\eta_1$  quite independent of  $\xi_2$  and  $\eta_2$ , we must have

$$\xi_1 \eta_1 - \eta_1 \xi_1 = i\hbar [\xi_1, \eta_1]$$

$$\xi_2 \eta_2 - \eta_2 \xi_2 = i\hbar [\xi_2, \eta_2],$$

where  $\hbar$  must not depend on  $\xi_1$  and  $\eta_1$ , nor on  $\xi_2$  and  $\eta_2$ , and also must commute with  $(\xi_1 \eta_1 - \eta_1 \xi_1)$ . It follows that  $\hbar$  must be simply a number. We want the P.B. of two observables or real variables to be real, as in the classical theory, which requires, from the work at the end of § 15, that  $\hbar$  shall be a real number when introduced, as here, with the coefficient  $i$ . We are thus led to the following *general formula for the quantum P.B.  $[\xi, \eta]$  of any two variables  $\xi$  and  $\eta$ ,*

$$\xi \eta - \eta \xi = i\hbar [\xi, \eta], \quad (7)$$

in which  $\hbar$  is a new universal constant having the dimensions of action. In order that the theory may agree with experiment, we must take  $\hbar$  equal to  $h/2\pi$ , where  $h$  is the universal constant that was introduced by Planck, known as Planck's constant. It is easily verified that the quantum P.B. satisfies all the conditions (2), (3), (4), (5) and (6).

The problem of finding quantum conditions now reduces to the problem of determining P.B.'s in quantum mechanics. The strong analogy between the quantum P.B. defined by (7) and the classical P.B. defined by (1) leads us to make the assumption that the quantum P.B.'s, or at any rate the simpler ones of them, have the same values as the corresponding classical P.B.'s. The simplest P.B.'s are those



involving the canonical coordinates and momenta themselves and have the following values in the classical theory:

$$\left. \begin{aligned} [q_r, q_s] &= 0 & [p_r, p_s] &= 0 \\ [q_r, p_s] &= \delta_{rs}. \end{aligned} \right\} (8)$$

We therefore assume that the corresponding quantum P.B.'s also have the values given by (8). By eliminating the quantum P.B.'s with the help of (7), we obtain the equations

$$\left. \begin{aligned} q_r q_s - q_s q_r &= 0 & p_r p_s - p_s p_r &= 0 \\ q_r p_s - p_s q_r &= i\hbar \delta_{rs}, \end{aligned} \right\} (9)$$

which are the *fundamental quantum conditions*. They show us where the lack of commutability among the canonical coordinates and momenta lies. They also provide us with a basis for calculating commutability relations between other dynamical variables. For instance, if  $\xi$  and  $\eta$  are any two functions of the  $q$ 's and  $p$ 's expressible as power series, we may express  $\xi\eta - \eta\xi$  or  $[\xi, \eta]$ , by repeated applications of the laws (2), (3), (4) and (5), in terms of the elementary P.B.'s given in (8) and so evaluate it. The result is often, in simple cases, the same as the classical result, or departs from the classical result only through requiring a special order for factors in a product, this order being, of course, unimportant in the classical theory. Even when  $\xi$  and  $\eta$  are more general functions of the  $q$ 's and  $p$ 's not expressible as power series, equations (9) are still sufficient to fix the value of  $\xi\eta - \eta\xi$ , as will become clear from the following work. Equations (9) thus give the solution of the problem of finding the quantum conditions, for all those dynamical systems which have a classical analogue and which are describable in terms of canonical coordinates and momenta. This does not include all possible systems in quantum mechanics.

Equations (7) and (9) provide the foundation for the analogy between quantum mechanics and classical mechanics. They show that *classical mechanics may be regarded as the limiting case of quantum mechanics when  $\hbar$  tends to zero*. A P.B. in quantum mechanics is a purely algebraic notion and is thus a rather more fundamental concept than a classical P.B., which can be defined only with reference to a set of canonical coordinates and momenta. For this reason canonical coordinates and momenta are of less importance in quantum mechanics than in classical mechanics; in fact, we may have a system in quantum mechanics for which canonical coordinates and momenta do



not exist and we can still give a meaning to P.B.'s. Such a system would be one without a classical analogue and we should not be able to obtain its quantum conditions by the method here described.

## 26. Canonical Coordinates and Momenta

Let us examine in greater detail the conditions (9) for canonical coordinates and momenta, which we assume to be all observables. One of the first things we notice is that two variables with different suffixes  $r$  and  $s$  always commute. It follows that any function of  $q_r$  and  $p_r$  will commute with any function of  $q_s$  and  $p_s$  when  $s$  differs from  $r$ . Thus *dynamical variables referring to different degrees of freedom commute*. This law, as we have derived it from (9), is proved only for dynamical systems with classical analogues, but we assume it to hold generally. In this way we can make a start on the problem of finding quantum conditions for dynamical systems for which canonical coordinates and momenta do not exist, in so far as we can give a meaning to degrees of freedom.

In applications of quantum mechanics it is often convenient to take two separate dynamical systems and to put them together and count them as forming a single system. This would be useful, for instance, if one wanted subsequently to introduce an interaction between the two systems and to treat this interaction perhaps by a perturbation method of the kind given in Chapter VIII. We can see from the above law how two dynamical systems may be counted as a single system. All the dynamical variables of one of the constituent systems must commute with all those of the other, since each of the two constituent systems has its own degrees of freedom. If we now take a complete set of commuting observables  $\xi_1$  for the first constituent system and a complete set of commuting observables  $\xi_2$  for the second, then it is easily seen that the  $\xi_1$ 's and  $\xi_2$ 's together form a complete set of commuting observables for the whole system; in fact, the basic  $\psi$ 's,  $\psi(\xi_1' \xi_2')$  in the  $(\xi_1 \xi_2)$ -representation for the whole system may each be considered as a product of the basic  $\psi$ 's,  $\psi(\xi_1')$  and  $\psi(\xi_2')$  in representations for the constituent systems, the  $\psi$ -space for the whole system being considered as the product of the  $\psi$ -spaces for the constituent systems. The product of representatives of  $\psi$ 's for the constituent systems will give the representative of a  $\psi$  for the whole system, thus,

$$(\xi_1' \xi_2') = (\xi_1')(\xi_2'), \quad (10)$$



although, of course, the general  $\psi$  for the whole system will not be of the form of the right-hand side of (10), but will be a sum or integral of terms of this form. If  $\eta_1$  and  $\eta_2$  denote a second pair of complete sets of commuting observables for the two constituent systems respectively, the transformation function for the whole system will be just the product of the transformation functions for the constituent systems, thus,

$$(\xi'_1 \xi'_2 | \eta'_1 \eta'_2) = (\xi'_1 | \eta'_1) (\xi'_2 | \eta'_2). \quad (11)$$

The multiplication laws (10) and (11) apply, of course, to any division of the degrees of freedom of the whole system into two sets, even though these two sets do not correspond physically to two constituent systems. The generalization to more than two constituent systems, or more than two sets of degrees of freedom, can be made immediately.

Let us now go back to the equations (9) and see what they tell us for a single degree of freedom. We have now just one  $q$  and one  $p$ , forming what is called in classical mechanics a pair of conjugate variables, and they satisfy

$$qp - pq = i\hbar. \quad (12)$$

Equation (12) is the fundamental equation in quantum mechanics for a pair of conjugate variables describing a degree of freedom that has a classical analogue. It is of such frequent occurrence that its main algebraic consequences should be noted.

We have

$$q^2p - pq^2 = q(qp - pq) + (qp - pq)q = 2i\hbar q.$$

The more general formula

$$q^n p - p q^n = n i \hbar q^{n-1} \quad (13)$$

is also valid. It is best proved by induction. Assuming (13) holds for one particular value of  $n$ , we find

$$\begin{aligned} q^{n+1}p - p q^{n+1} &= q(q^n p - p q^n) + (q p - p q) q^n \\ &= q \cdot n i \hbar q^{n-1} + i \hbar q^n = (n+1) i \hbar q^n, \end{aligned}$$

which is just (13) with  $n+1$  for  $n$ . Thus (13) holds generally. We may write (13) in the form

$$q^n p - p q^n = i \hbar \frac{d}{dq} q^n. \quad (14)$$

It follows that if  $f(q)$  is any function of  $q$  expressible as a power series,

$$f p - p f = i \hbar \frac{df}{dq}, \quad (15)$$



since we can apply (14) separately to each term in the expansion of  $f$ . In the next section at the end we shall see that (15) holds also for more general functions  $f$  that are not expressible as power series.

There is one example of (15) of special interest, namely when  $f$  is the exponential function  $e^{icq}$ ,  $c$  being any real number. This exponential function is defined in the usual way as the sum of the series

$$e^{icq} = \sum_{n=0}^{\infty} \frac{(ic)^n q^n}{n!} \quad (16)$$

and the ordinary exponential theorem must be valid for it, since there are no non-commuting quantities occurring in (16) to make a difference from ordinary algebra. With this expression for  $f$ , (15) becomes

$$e^{icq}p - pe^{icq} = -c\hbar e^{icq}$$

or

$$pe^{icq} = e^{icq}(p + c\hbar). \quad (17)$$

Let  $\psi_{p'}$  be an eigen- $\psi$  of  $p$  belonging to the eigenvalue  $p'$ , so that

$$p\psi_{p'} = p'\psi_{p'}.$$

From (17) we obtain

$$\begin{aligned} pe^{icq}\psi_{p'} &= e^{icq}(p + c\hbar)\psi_{p'} = e^{icq}(p' + c\hbar)\psi_{p'} \\ &= (p' + c\hbar)e^{icq}\psi_{p'}. \end{aligned} \quad (18)$$

Thus  $e^{icq}\psi_{p'}$  is an eigen- $\psi$  of  $p$  belonging to the eigenvalue  $p' + c\hbar$ . In this way we see that if  $p'$  is any eigenvalue of  $p$ ,  $p' + c\hbar$  must be another. Since  $c$  is arbitrary, it follows that the eigenvalues of  $p$  must include all numbers from  $-\infty$  to  $\infty$ . Similarly it may be shown that the eigenvalues of  $q$  include all numbers from  $-\infty$  to  $\infty$ . Hence *canonical variables satisfying (12) or (9) have as eigenvalues all numbers from  $-\infty$  to  $\infty$* . This result is known to be true from physical grounds in the case when the canonical variables are Cartesian coordinates and momenta of particles.

A possible source of difficulty in the above deduction should be pointed out. We could take  $c$  to be a pure imaginary or complex number and could then still formally deduce (17) and then (18). This would seem to lead to the result that  $p$  has complex numbers as eigenvalues, whereas  $p$  being, as we assumed at the beginning of this section, an observable, can have only real eigenvalues. The solution of the paradox lies in the fact that, for imaginary or complex  $c$ , the series (16) must be counted as non-convergent and the operator  $e^{icq}$  as not existing according to our general theory of linear operators given in Chapters II to IV. The eigenvalues of  $q$  extend to  $-\infty$  and to  $\infty$ , and



at one or other of these places  $e^{icq}$  would tend to infinity so rapidly as to be physically inadmissible as an operator, that can operate on  $\psi$ 's representing states that actually exist to give other  $\psi$ 's representing states that actually exist.

## 27. Momenta as Differential Operators

Let us take a dynamical system described by a set of canonical coordinates and momenta  $q_r, p_r$ , and introduce a representation in which the coordinates  $q_r$  are all diagonal. We may assume the  $q_r$  to form a complete set of commuting observables, the justification for this assumption being that it leads, as we shall see, to a self-consistent scheme of representatives for the  $q$ 's and  $p$ 's satisfying all the conditions (9). The representative of any  $\psi$  will thus be of the form  $(q'_1 q'_2 \dots q'_n |)$ . The domain of each of the variables  $q'$  extends from  $-\infty$  to  $\infty$ .

To begin our investigation, let us suppose our dynamical system has only one degree of freedom, so that we have to deal with just one  $q$  and one  $p$ , satisfying (12), and the representative of a  $\psi$  is simply  $(q' |)$ . An interesting linear operator now presents itself for study, namely the operator of differentiation of any  $(q' |)$  with respect to  $q'$ . This linear operator can be applied to the representative of any  $\psi$  and will give a function of  $q'$  which may be regarded as the representative of another  $\psi$ . This linear operator may therefore be handled according to our general theory of Chapters II to IV. Let us denote it by  $\pi$  in symbolic notation. It may be defined by the condition that if

$$\begin{aligned} \psi &= \int \psi(q') dq' (q' |), \\ \text{then } \pi\psi &= \int \psi(q') dq' \frac{d}{dq'} (q' |). \end{aligned} \quad \left. \vphantom{\int \psi(q') dq' \frac{d}{dq'} (q' |)} \right\} (19)$$

Its representative is thus determined by

$$\int (q' | \pi | q'') dq'' (q'' |) = \frac{d}{dq'} (q' |)$$

$$\text{and is therefore} \quad (q' | \pi | q'') = \delta'(q' - q''), \quad (20)$$

from (6) of Chapter IV. From (20) we can see that  $i\pi$  is Hermitian. We can also see that when  $\pi$  operates to the left on a  $\phi$ -symbol, it is equivalent to minus the operator of differentiation applied to the representative of that  $\phi$ -symbol; thus

$$\int (|q') dq' (q' | \pi | q'') = -\frac{d}{dq'} (|q'),$$



so that if

$$\phi = \int (|q'\rangle dq' \phi(q'),$$

then

$$\phi\pi = \int -\frac{d}{dq'}(|q'\rangle dq' \phi(q')). \quad \left. \vphantom{\int} \right\} (21)$$

Let us now work out the commutability relation connecting  $\pi$  with  $q$ . We have the equation

$$\frac{d}{dq'}\{q' \cdot (q'|)\} = q' \frac{d}{dq'}(q'|) + (q'|),$$

which, written in symbolic notation, becomes

$$\pi q\psi = q\pi\psi + \psi.$$

This equation holds for arbitrary  $\psi$  and we may therefore cancel out the factor  $\psi$ . We are then left with

$$\pi q - q\pi = 1, \quad (22)$$

which is the required commutability relation. It could have been obtained alternatively directly from the representative (20) with the help of properties of the  $\delta$  function given in § 21, in particular equation (10) there.

On comparing (22) with (12), we see that  $-i\hbar\pi$  satisfies the same commutability relation with  $q$  that  $p$  does. Their difference,  $p+i\hbar\pi$ , commutes with  $q$ . From the theorem at the top of page 60 and from our assumption that  $q$  by itself forms a complete set of commuting observables, it follows that  $p+i\hbar\pi$  must be a function of  $q$ , i.e.

$$p+i\hbar\pi = f(q). \quad (23)$$

Both  $p$  and  $i\hbar\pi$  are Hermitian operators and the function  $f$  is real. We shall now see that, by suitably choosing the phase factors in our  $q$ -representation, we can arrange to make  $f$  vanish and  $p$  just equal  $-i\hbar\pi$ .

Let us take a new representation in which  $q$  is diagonal, differing from our previous one in the phase factors of the basic  $\psi$ 's, and let us use stars to denote things referring to the new representation. The connexion between the basic  $\psi$ 's of the two representations will be of the form

$$\psi^*(q') = \psi(q')e^{i\gamma'}, \quad (24)$$

where  $\gamma' = \gamma(q')$  is any real function of  $q'$ . The new representation will give us a new operator  $\pi^*$ , defined, in a corresponding way to (19),



by the condition that if

$$\left. \begin{aligned} \psi &= \int \psi^*(q') dq' (q'|)^*, \\ \text{then } \pi^* \psi &= \int \psi^*(q') dq' \frac{d}{dq'} (q'|)^*. \end{aligned} \right\} (25)$$

Putting the same  $\psi$  on the left-hand sides of (19) and (25), we have

$$(q'|)^* = e^{-i\gamma'}(q'|)$$

$$\text{and so } \frac{d}{dq'} (q'|)^* = e^{-i\gamma'} \frac{d}{dq'} (q'|) - ie^{-i\gamma'} \frac{d\gamma'}{dq'} (q'|).$$

The second of equations (25) now gives us

$$\begin{aligned} \pi^* \psi &= \int \psi^*(q') dq' \left\{ e^{-i\gamma'} \frac{d}{dq'} (q'|) - ie^{-i\gamma'} \frac{d\gamma'}{dq'} (q'|) \right\} \\ &= \int \psi(q') dq' \left\{ \frac{d}{dq'} (q'|) - i \frac{d\gamma'}{dq'} (q'|) \right\} \\ &= \pi \psi - i \frac{d\gamma}{dq} \psi. \end{aligned}$$

Since this holds for arbitrary  $\psi$  we have

$$\pi^* = \pi - i \frac{d\gamma}{dq}.$$

$$\text{Hence } p + i\hbar\pi^* = p + i\hbar\pi + \hbar \frac{d\gamma}{dq} = f(q) + \hbar \frac{d\gamma}{dq} \quad (26)$$

from (23). We may now choose the function  $\gamma$ , which has been left arbitrary up to the present, so that the right-hand side of (26) vanishes. This will make  $p$  just equal to  $-i\hbar\pi^*$ .

We can easily extend this work to the case of  $n$  degrees of freedom. We then have  $n$  differentiation operators  $\pi_r$ , one for each degree of freedom, and we define them by replacing the second of equations (19) by

$$\pi_r \psi = \int \psi(q') dq' \frac{\partial}{\partial q'_r} (q'|). \quad (27)$$

The representative of  $\pi_r$  will be

$$\begin{aligned} (q'| \pi_r | q'') &= \delta(q'_1 - q''_1) \delta(q'_2 - q''_2) \dots \\ &\dots \delta(q'_{r-1} - q''_{r-1}) \delta'(q' - q'') \delta(q'_{r+1} - q''_{r+1}) \dots \delta(q'_n - q''_n), \end{aligned} \quad (28)$$

like the representative in the right-hand side of equation (34) of Chapter IV. From the form of this representative we again see that  $i\pi_r$  is Hermitian and that, when  $\pi_r$  operates to the left on a  $\phi$ -symbol,



it is equivalent to minus the operator of differentiation with respect to  $q_r$  applied to the representative of that  $\phi$ -symbol.

To obtain the commutability relations for the  $\pi_r$ 's, we note first that

$$\frac{\partial^2}{\partial q_r' \partial q_s'} (q' |) = \frac{\partial^2}{\partial q_s' \partial q_r'} (q' |),$$

which, written in symbolic notation, gives us

$$\pi_r \pi_s \psi = \pi_s \pi_r \psi.$$

This holds for arbitrary  $\psi$  and hence

$$\pi_r \pi_s - \pi_s \pi_r = 0. \quad (29)$$

Again 
$$\frac{\partial}{\partial q_s'} \{q_r' (q' |)\} = q_r' \frac{\partial}{\partial q_s'} (q' |) + \delta_{rs} (q' |),$$

which, written in symbolic notation, gives us

$$\pi_s q_r \psi = q_r \pi_s \psi + \delta_{rs} \psi$$

and hence

$$q_r \pi_s - \pi_s q_r = -\delta_{rs}. \quad (30)$$

Comparing (29) and (30) with (9), we see that the operators  $-i\hbar\pi_r$  satisfy the same commutability relations with the  $q$ 's and with each other as the  $p_r$  do. We can get a generalization of (30) by noticing that, if  $f(q)$  is any function of the  $q$ 's,

$$\frac{\partial}{\partial q_s'} \{f(q') (q' |)\} = f(q') \frac{\partial}{\partial q_s'} (q' |) + \frac{\partial f(q')}{\partial q_s'} (q' |).$$

Written in symbolic notation, this gives us

$$\pi_s f \psi = f \pi_s \psi + \frac{\partial f}{\partial q_s} \psi$$

and hence

$$f \pi_s - \pi_s f = -\frac{\partial f}{\partial q_s}. \quad (31)$$

From (30) and the corresponding equations for  $p_s$  in (9), we see that each of the operators  $p_s + i\hbar\pi_s$  commutes with each of the  $q$ 's. It follows as before, from the theorem at the top of page 60, that each  $p_s + i\hbar\pi_s$  is a function of the  $q$ 's only, i.e.

$$p_s + i\hbar\pi_s = f_s(q), \quad (32)$$

$f_s(q)$  being a function of the  $q$ 's, necessarily real. Using (29) and the corresponding equation for  $p$ 's in (9), we now obtain

$$\begin{aligned} 0 &= p_r p_s - p_s p_r = (-i\hbar\pi_r + f_r)(-i\hbar\pi_s + f_s) - (-i\hbar\pi_s + f_s)(-i\hbar\pi_r + f_r) \\ &= -i\hbar(\pi_r f_s + f_r \pi_s - \pi_s f_r - f_s \pi_r), \end{aligned}$$

or

$$\pi_s f_r - f_r \pi_s = \pi_r f_s - f_s \pi_r.$$



This gives, with the help of (31),

$$\frac{\partial f_r}{\partial q_s} = \frac{\partial f_s}{\partial q_r},$$

which shows that the functions  $f_r$  are all of the form

$$f_r = \partial F / \partial q_r,$$

where  $F$  is a function of the  $q$ 's independent of  $r$ . (It may be taken to be a real function.) Equations (32) now become

$$p_s = -i\hbar\pi_s + \partial F / \partial q_s. \quad (33)$$

Let us now, as in the case of one degree of freedom, introduce a new representation in which the  $q$ 's are all diagonal, differing from the previous one in the phase factors of the basic  $\psi$ 's, and let us again use stars to denote things referring to the new representation. The connexion between the basic  $\psi$ 's of the two representations will again be of the form (24),  $\gamma'$  now being an arbitrary function of all the variables  $q'$ . The new differentiation operators  $\pi_r^*$  will be defined by equations (25), with the second of these equations replaced by

$$\pi_s^* \psi = \int \psi^*(q') dq' \frac{\partial}{\partial q'_s} (q' |)^*,$$

corresponding to (27). We now obtain, by similar analysis to that which led to (26), the result

$$\pi_s^* = \pi_s - i \frac{\partial \gamma}{\partial q_s}.$$

Comparing this with (33), we see that, if we take  $\gamma = -F/\hbar$ , each  $p_s$  becomes just equal to  $-i\hbar\pi_s^*$ .

We have now established the general result that, *by suitably choosing the phase factors in a representation in which the  $q$ 's are diagonal, we can make each of the momenta conjugate to the  $q$ 's take the form*

$$p_s = -i\hbar\pi_s, \quad (34)$$

$\pi_s$  being the operator which, operating to the right on a  $\psi$ -vector, is equivalent to differentiation of the representative of that  $\psi$ -vector with respect to  $q_s$ , and operating to the left on a  $\phi$ -vector, is equivalent to minus the differentiation of the representative of that  $\phi$ -vector with respect to  $q_s$ . The interpretation of  $\pi_s$  as a differentiation or as minus a differentiation, according to whether it is multiplied to the right or to the left, is easily seen to apply generally, also when the thing it is multiplied into is not a  $\psi$  or a  $\phi$ . Thus, for example, with  $\pi_s$  multiplied



into a linear operator  $\xi$ , we have

$$(q'|\pi_s\xi|q'') = \frac{\partial}{\partial q'_s}(q'|\xi|q'')$$

$$(q'|\xi\pi_s|q'') = -\frac{\partial}{\partial q''_s}(q'|\xi|q''),$$

giving

$$\left. \begin{aligned} (q'|p_s\xi|q'') &= -i\hbar \frac{\partial}{\partial q'_s}(q'|\xi|q'') \\ (q'|\xi p_s|q'') &= i\hbar \frac{\partial}{\partial q''_s}(q'|\xi|q''). \end{aligned} \right\} \quad (35)$$

The result (34) is a very valuable one in applications of quantum mechanics. It is a consequence only of the quantum conditions (9) and may be regarded as a new way of expressing these conditions. We may illustrate its value by taking any function  $H(q_r, p_s)$  of the  $q$ 's and  $p$ 's expressible as a power series in the  $p$ 's. This function must be equal to  $H(q_r, -i\hbar\pi_s)$  and therefore, as an operator operating to the right on a  $\psi$ -vector, it must be equivalent to the differential operator

$$H\left(q_r, -i\hbar \frac{\partial}{\partial q'_s}\right), \quad (36)$$

in which each  $\pi_s$  has been replaced by  $\partial/\partial q'_s$  without any alteration of the order of factors in products, operating on the representative  $(q'|)$  of this  $\psi$ -vector. Thus  $H$  becomes expressed as a familiar kind of differential operator. The equation for determining the eigenvalues of  $H$  is

$$H\left(q_r, -i\hbar \frac{\partial}{\partial q'_s}\right)(q'|) = H' \cdot (q'|), \quad (37)$$

which is just a partial differential equation for the unknown function  $(q'|)$  and the unknown eigenvalue  $H'$ . A solution  $(q'|)$  of an equation like (37) is called an *eigenfunction* of the relevant operator  $H$ . Equations of the form (37) were introduced into quantum mechanics by *Schrödinger*.

We can now understand rather better the meaning of the indeterminacy in a representation when only the observables that are to be diagonal in it are specified, at any rate for the case when these observables are a set of canonical coordinates. Corresponding to each representation in which the  $q$ 's are diagonal there exists one set of momenta conjugate to the  $q$ 's [i.e. satisfying the same conditions as the  $p$ 's in (9)] whose representatives are of the specially simple form (34) instead of the more general form (33). If we take some particular



set of momenta conjugate to the  $q$ 's and require that these shall have representatives of the specially simple form, the representation is then completely determined, except for a trivial phase factor  $e^{i\beta}$  where  $\beta$  is independent of the  $q$ 's, since the function  $\gamma$  above is completely determined by the condition that each  $-i\hbar\pi_s$  shall equal  $p_s$ , except for an arbitrary constant.

As a corollary to the above work we may note that, from (31) applied in a representation in which (34) holds,

$$fp_s - p_sf = i\hbar \frac{\partial f}{\partial q_s}. \quad (38)$$

This is the generalization of equation (15) to the case of several degrees of freedom and the case when  $f$  is a function of the  $q$ 's not necessarily expressible as a power series.

## 28. Heisenberg's Principle of Uncertainty

On account of the general symmetry between the  $q$ 's and the  $p$ 's in the quantum conditions (9), it must be possible to interchange  $q$ 's and  $p$ 's throughout the work of the preceding section. This would mean setting up a representation in which the  $p$ 's are diagonal and each  $q$  is represented by the operator,  $\pm i\hbar$  times differentiation with respect to the corresponding  $p$ , the  $+$  sign being taken when it operates to the right and the  $-$  sign when it operates to the left. (These signs are just the other way round to what we had in the preceding section.) The two representations would be equally fundamental from the point of view of general theory. In practice, the representation in which the  $q$ 's are diagonal is the more useful one in general, since most of the dynamical quantities one has to deal with are expressible as power series in the  $p$ 's (usually of degree not higher than two), but are not expressible as power series in the  $q$ 's, and so would take the form of differential operators like (36) in the  $q$ -representation, but would not take the form of differential operators in the  $p$ -representation. There are, however, some problems in which the  $p$ -representation can be used with advantage, and it becomes desirable to calculate the transformation function connecting the two representations.

Let us take the case of the system with one degree of freedom and calculate the transformation function  $(q'|p')$  connecting the representation in which  $q$  is diagonal with that in which  $p$  is diagonal. We shall use the general method described at the end of § 18. Equation



(52) of that section, applied to our present problem, reads

$$\int (q' | p | q'') dq'' (q'' | p') = (q' | p') p'.$$

We can evaluate the left-hand side of this equation according to (34), provided the phase factors of the  $q$ -representation are suitably chosen. This gives

$$-i\hbar \frac{d}{dq'} (q' | p') = p' (q' | p').$$

The solution of this differential equation for  $(q' | p')$  is

$$(q' | p') = a' e^{iq'p'/\hbar}, \quad (39)$$

where  $a' = a(p')$  is an arbitrary function of  $p'$ . Note that (39) gives the general form of the  $q$ -representative of an eigen- $\psi$  of  $p$ .

We can determine the modulus of  $a'$  by using the normalizing condition

$$\int_{-\infty}^{\infty} (p' | q') dq' (q' | p'') = \delta(p' - p''),$$

which comes from (30) of Chapter IV. This gives, when we put

$$(p' | q') = \overline{(q' | p')} = \bar{a}' e^{-iq'p'/\hbar},$$

the equation

$$\bar{a}' a'' \int_{-\infty}^{\infty} e^{iq'(p''-p')/\hbar} dq' = \delta(p' - p''),$$

where  $a'' = a(p'')$ . Integrating the left-hand side with the help of (15) of Chapter IV, we obtain

$$2\pi \bar{a}' a'' \delta\{(p'' - p')/\hbar\} = \delta(p' - p''),$$

and hence, from (8), (11), and (14) of Chapter IV,

$$2\pi \hbar \bar{a}' a' = 1.$$

Thus  $a'$  is of the form  $\hbar^{-1/2} e^{i\gamma'}$ , where  $\gamma'$  is some real function of  $p'$ , and hence

$$(q' | p') = \hbar^{-1/2} e^{i\gamma'} e^{iq'p'/\hbar}.$$

By suitably choosing the phase factors of the  $p$ -representation [those of the  $q$ -representation were chosen when we made use of (34), but those of the  $p$ -representation are still arbitrary] we may remove the factor  $e^{i\gamma'}$ , leaving as final result

$$(q' | p') = \hbar^{-1/2} e^{iq'p'/\hbar}. \quad (40)$$

The result (40) shows that the formulas connecting the  $q$ - and



$p$ -representatives of a  $\psi$ -vector are

$$\left. \begin{aligned} (p'|) &= h^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-iq'p'/h} dq' (q'|), \\ (q'|) &= h^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{iq'p'/h} dp' (p'|). \end{aligned} \right\} (41)$$

These formulas have an elementary significance. They show that *either of the representatives is given, apart from numerical coefficients, by the amplitudes of the Fourier components of the other.*

It is interesting to apply (41) to a  $\psi$ -vector whose  $q$ -representative consists of what is called a *wave packet*. This is a function whose value is very small everywhere outside a certain domain, of width  $\Delta q'$  say, and inside this domain is approximately periodic with a definite frequency.† If a Fourier analysis is made of such a wave packet, the amplitude of all the Fourier components will be small, except those in the neighbourhood of the definite frequency. The components whose amplitudes are not small will fill up a frequency† band whose width is of the order  $1/\Delta q'$ , since two components whose frequencies differ by this amount, if in phase in the middle of the domain  $\Delta q'$ , will be just out of phase and interfering at the ends of this domain. Now in the first of equations (41) the variable  $p'/h$  plays the part of frequency. Thus with  $(q'|)$  of the form of a wave packet, the function  $(p'|)$ , being composed of the amplitudes of the Fourier components of the wave packet, will be small everywhere in the  $p'$ -space outside a certain domain of width  $\Delta p' = h/\Delta q'$ .

Let us now apply the physical interpretation of the square of the modulus of the representative of a  $\psi$  as a probability. We find that our wave packet represents a state for which a measurement of  $q$  is almost certain to lead to a result lying in a domain of width  $\Delta q'$  and a measurement of  $p$  is almost certain to lead to a result lying in a domain of width  $\Delta p'$ . We may say that for this state  $q$  has a definite value with an error of order  $\Delta q'$  and  $p$  has a definite value with an error of order  $\Delta p'$ . The product of these two errors is

$$\Delta q' \Delta p' = h. \quad (42)$$

Thus the more accurately one of the variables  $q, p$  has a definite value, the less accurately the other has a definite value. In the limit

† Frequency here means reciprocal of wave-length.



when one of them is completely determined, the other is completely undetermined. This last result can be obtained more directly from the transformation function  $(q'|p')$ . According to the end of § 22,  $|(q'|p')|^2 dq'$  is proportional to the probability of  $q$  having a value in the small range from  $q'$  to  $q'+dq'$  for the state for which  $p$  certainly has the value  $p'$ , and from (40) this probability is independent of  $q'$  for a given  $dq'$ . Thus if  $p$  certainly has a definite value  $p'$ , all values of  $q$  are equally probable. Similarly it may be shown that if  $q$  certainly has a definite value  $q'$ , all values of  $p$  are equally probable.

Equation (42) is known as *Heisenberg's Principle of Uncertainty*. It shows clearly the limitations in the possibility of simultaneously assigning numerical values, for any particular state, to two non-commuting observables, when those observables are canonically conjugate variables, and provides a plain illustration of how observations in quantum mechanics may be incompatible. It also shows how classical mechanics, which assumes that numerical values can be assigned simultaneously to all observables, may be a valid approximation when  $h$  can be considered as small enough to be negligible. Equation (42) holds only in the most favourable case, which occurs when the representative of the state is of the form of a wave packet. Other forms of representative would lead to a  $\Delta q'$  and  $\Delta p'$  whose product is larger than  $h$ .

The foregoing work can be easily extended to systems with several degrees of freedom. The transformation function connecting the  $q$ - and  $p$ -representations when there are  $n$  degrees of freedom is, according to the law (11), just the product of the transformation functions for each degree of freedom separately, namely

$$\begin{aligned} (q'_1 q'_2 \dots q'_n | p'_1 p'_2 \dots p'_n) &= (q'_1 | p'_1) (q'_2 | p'_2) \dots (q'_n | p'_n) \\ &= h^{-n/2} e^{i(q'_1 p'_1 + q'_2 p'_2 + \dots + q'_n p'_n)/h}. \end{aligned} \quad (43)$$

The idea of a wave packet can be extended to the case of several  $q$ 's, the function  $(q'|)$  having to be very small everywhere outside a certain domain of the  $q'$ -space and approximately periodic in each of the  $q'$ 's inside this domain. The principle of uncertainty then applies to each degree of freedom separately.

## 29. Displacement Operators

An instructive way of looking at some of the quantum conditions is provided by a study of displacement operators. These appear in the theory when we take into consideration that the scheme of rela-



tions between states and observables given in Chapter II is essentially a *physical* scheme, so that if certain states and observables are connected by some relation, on our displacing them all in a definite way (for example, displacing them all through a distance  $\delta x$  in the direction of the  $x$ -axis of Cartesian coordinates), the new states and observables would have to be connected by the same relation.

The displacement of a state or observable is a perfectly definite process physically. Thus to displace a state or observable through a distance  $\delta x$  in the direction of the  $x$ -axis, we should merely have to displace all the apparatus used in preparing the state, or all the apparatus required to measure the observable, through the distance  $\delta x$  in the direction of the  $x$ -axis, and the displaced apparatus would define the displaced state or observable. A displaced state or observable is uniquely determined by the undisplaced state or observable together with the direction and magnitude of the displacement.

The displacement of a  $\psi$ -vector is not such a definite thing though. If we take a certain  $\psi$ -vector, it will represent a certain state and we may displace this state and get a perfectly definite new state, but this new state will not determine our displaced  $\psi$ , but only the direction of our displaced  $\psi$ . We help to fix our displaced  $\psi$  by requiring that it shall have the same length as the undisplaced  $\psi$ , but even then it is not completely determined, but can still be multiplied by an arbitrary phase factor. One would think at first sight that each  $\psi$  one displaces would have a different independent phase factor, but with the help of the following extra condition, we see that they must all have the same. We here make use of the law that superposition relationships between states remain invariant under the displacement. A superposition relationship between states is expressed mathematically by a linear equation between the  $\psi$ 's representing those states, for example

$$\psi_0 = c_1 \psi_1 + c_2 \psi_2, \quad (44)$$

where  $c_1$  and  $c_2$  are numbers, and the invariance of the superposition relationship requires that the displaced states can be represented by  $\psi$ 's with the same linear equation between them—in our example they could be represented by  $\psi_0^\dagger$ ,  $\psi_1^\dagger$ , and  $\psi_2^\dagger$  satisfying

$$\psi_0^\dagger = c_1 \psi_1^\dagger + c_2 \psi_2^\dagger. \quad (45)$$

We now take such  $\psi$ 's to be our displaced  $\psi$ 's, that is to say, we require that *any linear equations holding between our undisplaced  $\psi$ 's*



shall hold also between our displaced  $\psi$ 's. This makes it impossible to provide our displaced  $\psi$ 's with independently variable phase factors, as these would spoil the linear equations [for example, (45) would cease to be valid if we multiplied  $\psi_0^\dagger$ ,  $\psi_1^\dagger$ , and  $\psi_2^\dagger$  by different factors  $e^{i\gamma_0}$ ,  $e^{i\gamma_1}$ , and  $e^{i\gamma_2}$ ], and the only arbitrariness left in the displaced  $\psi$ 's is that of a single arbitrary phase factor to be multiplied into them all.

With the displacement of  $\psi$ 's made fairly definite in the above manner and the displacement of  $\phi$ 's, of course, made equally definite, through their being the conjugate imaginaries of the  $\psi$ 's, we can now assert that any symbolic equation between  $\psi$ 's,  $\phi$ 's, and linear operators must remain invariant under the displacement of every symbol occurring in it, on account of such an equation having some physical significance which will not get changed by the displacement. Take, for example, the equation

$$\phi_k \psi_l = c, \quad (46)$$

$\phi_k$  being any  $\phi$ -vector,  $\psi_l$  any  $\psi$ -vector, and  $c$  a number, equal to their scalar product. The assertion that this equation remains invariant under the displacement may be written, if we use the sign  $\dagger$  generally to denote a displaced quantity,

$$\phi_k^\dagger \psi_l^\dagger = c = \phi_k \psi_l, \quad (47)$$

and is thus equivalent to the assertion that the scalar product  $\phi_k \psi_l$  is invariant. Now a scalar product  $\phi_k \psi_l$  may be regarded as a specification of the extent to which the two states represented by  $\phi_k$  and  $\psi_l$  approximate to being orthogonal (it vanishes when they are orthogonal), and the assertion of its invariance is justified on account of the notion of orthogonality of two states being a physical notion, unaffected by an equal displacement of both states. Again, an equation of the type

$$\xi \psi_a = \psi_b, \quad (48)$$

$\xi$  being any observable, denotes some physical relation between the observable  $\xi$  and the two states represented by  $\psi_a$  and  $\psi_b$ , although this relation cannot be described in an elementary way. This physical relation must be invariant under the displacement and hence equation (48) must be invariant.

To deal mathematically with the invariance of equations like (46) and (48), it is convenient to introduce a process of differentiation,



denoted by  $D_x$ , defined by

$$\left. \begin{aligned} D_x \psi_l &= \lim_{\delta x \rightarrow 0} \frac{\psi_l^\dagger - \psi_l}{\delta x} \\ D_x \phi_k &= \lim_{\delta x \rightarrow 0} \frac{\phi_k^\dagger - \phi_k}{\delta x} \\ D_x \xi &= \lim_{\delta x \rightarrow 0} \frac{\xi^\dagger - \xi}{\delta x}, \end{aligned} \right\} \quad (49)$$

the  $\dagger$  denoting a quantity displaced through a distance  $\delta x$  in the direction of the  $x$ -axis. There will be some lack of determinacy in  $D_x \psi_l$  due to the arbitrary phase factor by which we may multiply all our displaced  $\psi$ 's. Taking new displaced  $\psi$ 's equal to  $e^{i\gamma}$  times the previous ones, we get a new  $D_x \psi_l$ , say  $D_x^* \psi_l$ , defined by

$$\begin{aligned} D_x^* \psi_l &= \lim_{\delta x \rightarrow 0} \frac{e^{i\gamma} \psi_l^\dagger - \psi_l}{\delta x} \\ &= \lim_{\delta x \rightarrow 0} \left\{ \frac{\psi_l^\dagger - \psi_l}{\delta x} + \frac{e^{i\gamma} - 1}{\delta x} \psi_l^\dagger \right\} \\ &= D_x \psi_l + ia \psi_l, \end{aligned} \quad (50)$$

where  $a$  is a real number and is the limit of  $\gamma/\delta x$ . (We must choose  $\gamma$  so that this limit exists in order that  $D^*$  may have a meaning.) There will be a corresponding lack of determinacy in  $D_x \phi_k$ , but none in  $D_x \xi$ . Applying our differentiation process  $D_x$ , which is subject to the usual law for the differentiation of a product, to equations (46) and (48), we get

$$(D_x \phi_k) \psi_l + \phi_k (D_x \psi_l) = 0 \quad (51)$$

$$\text{and} \quad (D_x \xi) \psi_a + \xi (D_x \psi_a) = D_x \psi_b. \quad (52)$$

These equations must hold for each of the various meanings of  $D_x$  arising from its lack of determinacy.

The condition that linear equations between the  $\psi$ 's remain invariant under the displacement and that an equation such as (45) holds whenever the corresponding (44) holds, means that the displaced  $\psi$ 's are linear functions of the undisplaced  $\psi$ 's and that each displaced  $\psi$  is the result of some linear operator $\dagger$  applied to the corresponding undisplaced  $\psi$ . In symbols,

$$\psi_l^\dagger = A \psi_l, \quad (53)$$

where  $A$  is a linear operator independent of  $\psi_l$  and depending only on

$\dagger$  This follows at once (with the definition of a linear operator given in § 8) from the invariance of the linear equation expressing an arbitrary  $\psi$  in terms of the basic  $\psi$ 's of a representation.



the displacement. It follows that  $D_x \psi_l$  must also be the result of some linear operator applied to  $\psi_l$ . We call this linear operator the displacement operator  $d_x$ , thus

$$D_x \psi_l = d_x \psi_l. \quad (54)$$

Alternatively, we could define  $d_x$  directly in terms of the operator  $A$  of equation (53), as

$$d_x = \lim_{\delta x \rightarrow 0} \frac{A - 1}{\delta x}. \quad (55)$$

From (50) we see that the lack of determinacy in  $d_x$  consists in the possibility of adding to it an arbitrary, pure imaginary number.

Let us see how to express  $D_x \phi_k$  and  $D_x \xi$  in terms of  $d_x$ ,  $\phi_k$ , and  $\xi$ . From (51) and (54) we get

$$(D_x \phi_k) \psi_l + \phi_k d_x \psi_l = 0,$$

and since this holds for arbitrary  $\psi_l$ , we must have

$$D_x \phi_k = -\phi_k d_x. \quad (56)$$

This result shows that  $d_x$  is a pure imaginary operator ( $i$  times a Hermitian operator), since,  $D_x \phi_k$  being the conjugate imaginary of  $D_x \psi_k$ , it gives  $-\phi_k d_x$  as the conjugate imaginary of  $d_x \psi_k$ . From (52) and (54) we get

$$(D_x \xi) \psi_a + \xi d_x \psi_a = d_x \psi_b = d_x \xi \psi_a$$

from (48). Since this holds for arbitrary  $\psi_a$ , we must have

$$D_x \xi = d_x \xi - \xi d_x. \quad (57)$$

We can see from this result how it is that the lack of determinacy in  $d_x$ , consisting in the possibility of adding to it an arbitrary, pure imaginary number, is not associated with any lack of determinacy in  $D_x \xi$ .

Let us now introduce a set of canonical coordinates and momenta consisting of  $x$ ,  $y$ , and  $z$ , the Cartesian coordinates of the centre of gravity of our system, and  $p_x$ ,  $p_y$ , and  $p_z$ , the components of the total momentum of the system, which are the conjugates of  $x$ ,  $y$ , and  $z$ , together with any other coordinates and momenta that may be necessary for describing internal degrees of freedom of the system. If we suppose a piece of apparatus, which has been set up to measure  $x$ , to be displaced a distance  $\delta x$  in the direction of the  $x$ -axis, it will measure  $x - \delta x$ . Thus

$$x^\dagger = x - \delta x$$

and therefore, from the third of equations (49),

$$D_x x = -1. \quad (58)$$

From (57) we now find

$$x d_x - d_x x = 1.$$



This is the quantum condition connecting  $d_x$  with  $x$ . From similar arguments we find that each of the other canonical coordinates and momenta introduced above, since it is unaffected by the displacement, must commute with  $d_x$ . Comparing these results with (9), we see that  $i\hbar d_x$  satisfies just the same quantum conditions as  $p_x$ . Their difference,  $p_x - i\hbar d_x$ , commutes with all the coordinates and momenta and must therefore be a number. This number, which is necessarily real since  $p_x$  and  $i\hbar d_x$  are both Hermitian operators, may be made zero by a suitable choice of the arbitrary, pure imaginary number that can be added to  $d_x$ . We then have the result

$$p_x = i\hbar d_x, \quad (59)$$

or the  $x$ -component of the total momentum of the system is  $i\hbar$  times the displacement operator  $d_x$ .

This is a fundamental result, which gives a new significance to displacement operators. There is a corresponding result, of course, also for the  $y$  and  $z$  displacement operators  $d_y$  and  $d_z$ . The quantum conditions which state that  $p_x$ ,  $p_y$ , and  $p_z$  commute with each other are now seen to be connected with the fact that displacements in different directions are commutable operations.

We can build up a similar theory for rotation operators about the  $x$ ,  $y$  and  $z$  axes. These linear operators,  $d_\xi$ ,  $d_\eta$  and  $d_\zeta$  say, are found, in the same way as  $d_x$ , to be pure imaginary and to be undetermined to the extent of arbitrary, pure imaginary additive numbers. Their quantum conditions may be easily calculated and turn out to be, apart from the factor  $i\hbar$ , the same as those for the components of angular momentum of the system (as they will be deduced in § 38), so that we can identify  $i\hbar d_\xi$ ,  $i\hbar d_\eta$ , and  $i\hbar d_\zeta$  with the components of angular momentum. An interesting consequence of this result is that if a state, represented by  $\psi$ , has zero angular momentum, then

$$d_\xi \psi = d_\eta \psi = d_\zeta \psi = 0,$$

which requires that  $\psi$  shall be spherically symmetrical. Thus a state of zero angular momentum is necessarily spherically symmetrical.

### 30. Contact Transformations

Let  $U$  be any linear operator that has a reciprocal  $U^{-1}$  and consider the equation

$$\alpha^* = U\alpha U^{-1}, \quad (60)$$

$\alpha$  being an arbitrary linear operator. This equation may be regarded as expressing a transformation from any linear operator  $\alpha$  to a



corresponding linear operator  $\alpha^*$ , and as such it has rather remarkable properties. In the first place it should be noted that each  $\alpha^*$  has the same eigenvalues as the corresponding  $\alpha$ ; since, if  $\alpha'$  is any eigenvalue of  $\alpha$  and  $\psi_{\alpha'}$  is the eigen- $\psi$  belonging to it, we have

$$\alpha\psi_{\alpha'} = \alpha'\psi_{\alpha'}$$

and hence  $\alpha^*U\psi_{\alpha'} = U\alpha U^{-1}U\psi_{\alpha'} = U\alpha\psi_{\alpha'} = \alpha'U\psi_{\alpha'}$ ,

showing that  $U\psi_{\alpha'}$  is an eigen- $\psi$  of  $\alpha^*$  belonging to the same eigenvalue  $\alpha'$ , and similarly any eigenvalue of  $\alpha^*$  may be shown to be also an eigenvalue of  $\alpha$ . Further, if we take several  $\alpha$ 's that are connected by algebraic equations and transform them all according to (60), the corresponding  $\alpha^*$ 's will be connected by the same algebraic equations. This result follows from the fact that the fundamental algebraic processes of addition and multiplication are left invariant by the transformation (60), as is shown by the following equations:

$$(\alpha_1 + \alpha_2)^* = U(\alpha_1 + \alpha_2)U^{-1} = U\alpha_1 U^{-1} + U\alpha_2 U^{-1} = \alpha_1^* + \alpha_2^*$$

$$(\alpha_1 \alpha_2)^* = U\alpha_1 \alpha_2 U^{-1} = U\alpha_1 U^{-1}U\alpha_2 U^{-1} = \alpha_1^* \alpha_2^*.$$

Let us now see what condition would be imposed on  $U$  by the requirement that any Hermitian  $\alpha$  shall be transformed into a Hermitian  $\alpha^*$ . Equation (60) may be written

$$\alpha^*U = U\alpha. \quad (61)$$

Taking the conjugate complex of both sides in accordance with (19) of Chapter III we find, if  $\alpha$  and  $\alpha^*$  are both Hermitian,

$$\bar{U}\alpha^* = \alpha\bar{U}. \quad (62)$$

Equation (61) gives us  $\bar{U}\alpha^*U = \bar{U}U\alpha$

and equation (62) gives us

$$\bar{U}\alpha^*U = \alpha\bar{U}U.$$

Hence

$$\bar{U}U\alpha = \alpha\bar{U}U.$$

Thus  $\bar{U}U$  commutes with any Hermitian operator and therefore also with any linear operator whatever, since any linear operator can be expressed as one Hermitian operator plus  $i$  times another. It follows that  $\bar{U}U$  is a number. By taking a matrix representation we can easily see that this number must be real and positive. We can suppose it to be unity without any loss of generality in the transformation (60). We then have

$$\bar{U}U = 1. \quad (63)$$



Equation (63) is equivalent to any of the following

$$U = \bar{U}^{-1}, \quad \bar{U} = U^{-1}, \quad U^{-1}\bar{U}^{-1} = 1. \quad (64)$$

A matrix or linear operator  $U$  that satisfies (63) and (64) is said to be *unitary* and a transformation (60) with unitary  $U$  is called a *unitary transformation*. A unitary transformation transforms Hermitian operators into Hermitian operators. Also it transforms linear operators satisfying the expansion theorem into linear operators satisfying the expansion theorem, since, if  $\alpha$  satisfies the expansion theorem, we can expand  $U^{-1}\psi$ , where  $\psi$  is arbitrary, in terms of  $\psi_{\alpha'}$ 's and by multiplying this result by  $U$ , we get  $\psi$  expanded in terms of  $\psi$ -vectors of the form  $U\psi_{\alpha'}$ , each of which is an eigen- $\psi$  of  $\alpha^*$ . We can now see that a unitary transformation transforms observables into observables. It leaves invariant any algebraic equation between the observables and also, as may easily be verified, any functional relation based on the general definition of a function given in § 11.

The inverse of a unitary transformation is also a unitary transformation, owing to the fact, which follows from (64), that if  $U$  is unitary,  $U^{-1}$  is also unitary. Further, if two unitary transformations are applied in succession, the result is a third unitary transformation, as may be verified in the following way. Let the two unitary transformations be (60) and

$$\alpha^\dagger = V\alpha^*V^{-1}.$$

The connexion between  $\alpha^\dagger$  and  $\alpha$  is then

$$\begin{aligned} \alpha^\dagger &= VU\alpha U^{-1}V^{-1} \\ &= (VU)\alpha(VU)^{-1} \end{aligned} \quad (65)$$

from (32) of Chapter II. Now  $VU$  is unitary since

$$\overline{VUVU} = \bar{U}\bar{V}VU = \bar{U}U = 1,$$

and hence (65) is a unitary transformation.

A transformation from one set of canonical coordinates and momenta  $q_r, p_r$  to another set  $q_r^*, p_r^*$  is called in quantum mechanics, as in classical mechanics, a *contact transformation*. In quantum mechanics the conditions for a set of variables to be canonical are algebraic, namely equations (9), which makes the theory of contact transformations more elementary than in classical mechanics. We shall now see that quantum contact transformations are the same as the above unitary transformations.

Let us consider the contact transformation from the canonical



variables  $q_r, p_r$  to the canonical variables  $q_r^*, p_r^*$ . We shall use two representations in which the  $q$ 's and the  $q^*$ 's respectively are diagonal, the phase factors of these representations being such that equation (34) and the corresponding equation for the starred variables hold. We introduce the linear operator  $U$  whose mixed representative  $(q^{*'}|U|q'')$  is defined to be

$$(q^{*'}|U|q'') = \delta(q^{*'} - q''). \quad (66)$$

[The right-hand side of (66) has a meaning since each  $q'$  and  $q^{*'}$  takes on all values from  $-\infty$  to  $\infty$ .] We note in the first place that  $U$  is unitary, since, using fundamental equations of the transformation theory,<sup>†</sup>

$$(q'|U|q'') = \int (q'|q^{*''}) dq^{*''} (q^{*''}|U|q'') = (q'|q^{*''}),$$

so that

$$(q''|\bar{U}|q') = \overline{(q'|q^{*''})} = (q^{*''}|q'),$$

and hence

$$\begin{aligned} (q'|U\bar{U}|q''') &= \int (q'|U|q'') dq'' (q''|\bar{U}|q''') \\ &= \int (q'|q^{*''}) dq^{*''} (q^{*''}|q''') \\ &= \delta(q' - q'''), \end{aligned}$$

so that

$$U\bar{U} = 1.$$

We have further

$$(q^{*'}|q_r^* U|q'') = q_r^{*'} \delta(q^{*'} - q'')$$

and

$$(q^{*'}|U q_r|q'') = \delta(q^{*'} - q'') q_r''.$$

The right-hand sides of these two equations are equal on account of (9) of Chapter IV and hence

$$q_r^* U = U q_r$$

or

$$q_r^* = U q_r U^{-1}.$$

Again, according to the rules expressed by (35), which are valid also for mixed representatives

$$(q^{*'}|p_r^* U|q'') = -i\hbar \frac{\partial}{\partial q_r^{*'}} \delta(q^{*'} - q'')$$

$$(q^{*'}|U p_r|q'') = i\hbar \frac{\partial}{\partial q_r''} \delta(q^{*'} - q'').$$

<sup>†</sup> In this piece of analysis we use the notation that a  $q_r$  and a  $q_r^*$  with the same number of primes both denote the same number. Thus, for example,  $q_r'' = q_r^{*''}$ . It is necessary to retain both symbols for the same number in order to preserve the meaning of bracket expressions, such as  $(q'|q^{*''})$ .



The right-hand sides of these two equations are obviously equal, and hence

$$p_r^* U = U p_r$$

or

$$p_r^* = U p_r U^{-1}.$$

This establishes that a contact transformation is just a unitary transformation of the form (60). The converse result, that a unitary transformation applied to a set of canonical variables gives a contact transformation, is obvious, owing to the invariance of algebraic relations under a unitary transformation. We can now give a meaning to contact transformations for dynamical systems in which canonical coordinates and momenta do not exist, defining such transformations simply as unitary transformations.

One of the ways of expressing the conditions for a contact transformation in classical mechanics is

$$p_r = \frac{\partial S}{\partial q_r}, \quad p_r^* = -\frac{\partial S}{\partial q_r^*}, \quad (67)$$

$S$  being some function of the  $q$ 's and  $q^*$ 's. There is a quantum analogue of this. We define the quantum  $S$  by

$$(q' | q^{*''}) = e^{iS(q', q^{*''})/\hbar}. \quad (68)$$

We now have

$$\begin{aligned} (q' | p_r | q^{*''}) &= \int (q' | p_r | q''') dq''' (q''' | q^{*''}) \\ &= -i\hbar \frac{\partial}{\partial q_r'} (q' | q^{*''}) = \frac{\partial S(q', q^{*''})}{\partial q_r'} (q' | q^{*''}). \end{aligned} \quad (69)$$

Similarly,

$$\begin{aligned} (q' | p_r^* | q^{*''}) &= \int (q' | q^{*'''} ) dq^{*'''} (q^{*'''} | p_r | q^{*''}) \\ &= i\hbar \frac{\partial}{\partial q_r^{*''}} (q' | q^{*''}) = -\frac{\partial S(q', q^{*''})}{\partial q_r^{*''}} (q' | q^{*''}). \end{aligned} \quad (70)$$

From equation (50) of Chapter III we have

$$(q' | f(q)g(q^*) | q^{*''}) = f(q')g(q^{*''})(q' | q^{*''}), \quad (71)$$

where  $f(q)$  and  $g(q^*)$  are functions of the  $q$ 's and  $q^*$ 's respectively. Let  $B(q, q^*)$  be any function of the  $q$ 's and  $q^*$ 's consisting of a sum of terms each of the form  $f(q)g(q^*)$ , so that all the  $q$ 's in  $B$  occur to the left of all the  $q^*$ 's. Such a function we call *well-ordered*. Applying (71) to each of the terms in  $B$  and adding, we get

$$(q' | B(q, q^*) | q^{*''}) = B(q', q^{*''})(q' | q^{*''}). \quad (72)$$

Now let us suppose each  $p_r$  and  $p_r^*$  can be expressed as a well-ordered



function of the  $q$ 's and  $q^*$ 's and write these functions  $p_r(q, q^*)$ ,  $p_r^*(q, q^*)$ . We shall then have, from (72),

$$(q' | p_r | q^{*''}) = p_r(q', q^{*''})(q' | q^{*''}) \quad (73)$$

$$(q' | p_r^* | q^{*''}) = p_r^*(q', q^{*''})(q' | q^{*''}). \quad (74)$$

Comparing (73) with (69) and (74) with (70), we see that

$$p_r(q', q^{*''}) = \frac{\partial S(q', q^{*''})}{\partial q_r'}, \quad p_r^*(q', q^{*''}) = -\frac{\partial S(q', q^{*''})}{\partial q_r^{*''}}.$$

This means that

$$p_r = \frac{\partial S(q, q^*)}{\partial q_r}, \quad p_r^* = -\frac{\partial S(q, q^*)}{\partial q_r^*}, \quad (75)$$

provided the right-hand sides of (75) are written as well-ordered functions. Thus the classical equations (67) for a contact transformation hold also in the quantum theory when the non-commuting variables  $q$  and  $q^*$  in their right-hand sides are suitably ordered.

We get an infinitesimal contact or unitary transformation by taking  $U$  in (60) to differ by an infinitesimal from unity. Put

$$U = 1 + i\epsilon F,$$

where  $\epsilon$  is infinitesimal, so that its square can be neglected. Then

$$U^{-1} = 1 - i\epsilon F.$$

The unitary condition (63) or (64) requires that  $F$  shall be Hermitian. The transformation equation (60) now takes the form

$$\alpha^* = (1 + i\epsilon F)\alpha(1 - i\epsilon F),$$

$$\text{which gives} \quad \alpha^* - \alpha = i\epsilon(F\alpha - \alpha F). \quad (76)$$

It may be written in P.B. notation

$$\alpha^* - \alpha = \epsilon\hbar[\alpha, F],$$

when it is formally the same as a classical infinitesimal contact transformation.



## VI

### THE EQUATIONS OF MOTION

#### 31. Schrödinger's Form for the Equations of Motion

OUR work in Chapters II to V was all concerned with one instant of time. It gave the general scheme of relations between states and observations at that one instant of time. To get a complete theory of dynamics we must consider also the connexion between different instants of time and set up something of the nature of equations of motion.

The state of our system at each instant of time will be represented by some vector  $\psi$  and we have to find the law of variation of  $\psi$  with the time  $t$ . For this purpose we use the general principle of superposition, according to which, as discussed in § 6, any superposition relationship between states holding at one instant of time will hold throughout all time. Thus if, for example, we have three states at one instant of time, represented by three vectors  $\psi_0, \psi_1, \psi_2$  satisfying

$$\psi_0 = c_1 \psi_1 + c_2 \psi_2,$$

these states will vary with the time in such a way that at any other instant of time they will be represented by three vectors,  $\psi_0^\dagger, \psi_1^\dagger, \psi_2^\dagger$  say, which satisfy, provided the arbitrary numerical factors by which these vectors may be multiplied are suitably chosen,

$$\psi_0^\dagger = c_1 \psi_1^\dagger + c_2 \psi_2^\dagger,$$

with the same coefficients  $c_1$  and  $c_2$ . This requires, as we had in § 29 in connexion with equations (44) and (45) referring to a displacement, that each  $\psi^\dagger$  shall be the result of some linear operator applied to the corresponding  $\psi$ . If we now take the second instant of time, to which  $\psi^\dagger$  belongs, to differ by only a small time interval  $\delta t$  from the first and form the differential coefficient

$$\frac{d\psi}{dt} = \lim_{\delta t \rightarrow 0} \frac{\psi^\dagger - \psi}{\delta t},$$

then  $d\psi/dt$  must also be the result of some linear operator applied to the corresponding  $\psi$ .

We put

$$i\hbar \frac{d\psi}{dt} = H\psi, \tag{1}$$

where  $H$  is a linear operator independent of  $\psi$ . This gives the general



law for the variation of  $\psi$ -vectors with the time. We make the further assumption that  $H$  is a Hermitian operator. This has the effect of making any scalar product of a  $\phi$  with a  $\psi$  constant, since it causes the conjugate imaginary of (1) to be

$$-i\hbar \frac{d\phi}{dt} = \phi H, \quad (2)$$

so that

$$i\hbar \frac{d}{dt}(\phi_k \psi_l) = i\hbar \frac{d\phi_k}{dt} \psi_l + i\hbar \phi_k \frac{d\psi_l}{dt} = -(\phi_k H) \psi_l + \phi_k (H \psi_l) = 0.$$

By arguments similar to those used for  $d_x$  in § 29, we can deduce that  $H$  is undetermined to the extent of an arbitrary real additive number.

Formula (1) shows how all the states of our system vary with the time and is one of the fundamental ways of expressing the equations of motion of quantum mechanics. Written in terms of representatives in a representation in which, say, each of the complete set of commuting observables  $q$  is diagonal, it appears as†

$$i\hbar \frac{d}{dt}(q'|) = \int (q'|H|q'') dq'' (q''|). \quad (3)$$

In this form it is known as *Schrödinger's wave equation*, having been first put forward by Schrödinger in 1926, and is very extensively used in practical applications of the theory. Its solutions are called *wave functions*, owing to the fact that in a great many problems they are of the kind of function which represents waves; in fact, as we shall see in § 34, they are so, if the  $q$ 's are taken to be dynamical coordinates, in all those problems in which the classical theory holds as an approximation. The square of the modulus of a normalized solution gives the probability of the  $q$ 's having specified values at any time for some particular state of motion of the system. Formula (2) written in terms of representatives gives the conjugate complex equation to (3), namely

$$-i\hbar \frac{d}{dt}(|q'|) = \int (|q''| dq'' (q''|H|q'), \quad (4)$$

which is equally fundamental in general theory but is not so often explicitly used in practice.

The linear operator  $H$  introduced in (1) we call the *Hamiltonian* of the system. There is one such linear operator for each dynamical

† The case of continuous  $q$ 's is taken for definiteness, the usual modifications of notation being required for the discrete case.



system. We assume it to be always an observable and to be, in fact, the total energy of the system. Its analogy with the Hamiltonian of classical mechanics will become apparent in the next section. Like the classical Hamiltonian it may either be constant or vary with the time, one or other of these possibilities occurring according to whether there are present only forces of interaction between the various component parts of the system or whether there are also external forces present. The constancy or variability with time of the linear operator  $H$  implies, of course, the constancy or variability with time of its representative  $(q'|H|q'')$ .

When  $H$  is constant we can write down a formal solution of (1), namely

$$\psi_t = e^{-iHt/\hbar}\psi_0, \quad (5)$$

$\psi_0$  being the value of any  $\psi$  at time 0 and  $\psi_t$  its value at time  $t$ . This solution may be verified by direct substitution in (1), it being noted that the differentiation of the exponential can be carried out in the ordinary way since there are no non-commuting quantities involved. In practical problems the solution (5) is not often of use, owing to the difficulty of evaluating the exponential, and one usually has to work from the differential equation (3) instead.

It may happen that a particular state of our system does not vary with the time. It is then called a *stationary state*. The condition for a state to be stationary is that it shall be represented by a  $\psi$  whose direction remains constant, i.e.

$$\frac{d\psi}{dt} = \lambda\psi, \quad (6)$$

where  $\lambda$  is a number. Combining this equation with (1) we get

$$H\psi = i\hbar\lambda\psi,$$

which is just the condition that  $\psi$  shall be an eigen- $\psi$  of  $H$ . Thus *the stationary states are the eigenstates of the Hamiltonian*. It is necessary that equation (6) shall hold throughout all time and hence  $\psi$  must be an eigen- $\psi$  of  $H$  throughout all time. This is usually possible only when  $H$  is constant, so that stationary states usually exist only for a dynamical system with constant Hamiltonian. There are then so many of them that an arbitrary state is dependent on them (from our assumption that the Hamiltonian is an observable). For each of these stationary states the Hamiltonian or energy has a definite value, namely the eigenvalue  $H'$  to which the state belongs, equal to



$i\hbar$  times the  $\lambda$  of equation (6), and the  $\psi$  representing the state varies with time according to the law

$$\psi_t = e^{-iHt/\hbar}\psi_0, \quad (7)$$

i.e. the simple harmonic law, with a frequency depending only on the associated energy value.

### 32. Heisenberg's Form for the Equations of Motion

In the preceding section we had a picture of the states of our dynamical system represented by vectors in a certain vector space, these vectors varying with time in order to correspond to the changes taking place in the states. We shall call this the *Schrödinger picture*. On account of the linear form of the law of variation of the vectors with time, as shown by equation (1), we may adopt an alternative picture in which the vectors representing the states are all fixed, but are referred to a moving system of coordinates. We shall call this the *Heisenberg picture*. The two pictures are, of course, formally equivalent. In both of them the coordinates of a  $\psi$  representing a state vary in the same way, namely according to (3), the only difference being that in one of them this variation is ascribed to a motion of the  $\psi$ 's themselves and in the other it is ascribed to a motion of the system of coordinates.

In the Schrödinger picture a dynamical variable is represented by a constant linear operator. In the Heisenberg picture a dynamical variable will be represented by a linear operator fixed relative to the coordinate system and therefore, in general, varying with time. Let us determine its law of variation.

A vector  $\psi$  fixed relative to the coordinate system in the Heisenberg picture must vary with time according to the formula

$$i\hbar \frac{d\psi}{dt} = -H\psi, \quad (8)$$

that is, formula (1) with a minus sign, since this is the time-variation which must be superposed on (1) to bring  $\psi$  to rest. The  $H$  in (8) is at any time the same function of the dynamical variables as the  $H$  in (1), though these dynamical variables are now represented by moving linear operators. The condition for a linear operator  $\xi$  to be fixed relative to the coordinate system is that, when multiplied into any vector  $\psi_a$  fixed relative to the coordinate system, the resulting  $\psi$ -vector

$$\xi\psi_a = \psi_b \quad (9)$$



shall also be fixed relative to the coordinate system. Differentiating (9), we get

$$\frac{d\xi}{dt}\psi_a + \xi\frac{d\psi_a}{dt} = \frac{d\psi_b}{dt},$$

and with the help of formula (8) applied to  $\psi_a$  and  $\psi_b$ , we find

$$i\hbar\frac{d\xi}{dt}\psi_a - \xi H\psi_a = -H\psi_b = -H\xi\psi_a.$$

Since this holds for arbitrary  $\psi_a$  we can cancel out  $\psi_a$ , obtaining

$$i\hbar\frac{d\xi}{dt} = \xi H - H\xi. \quad (10)$$

Equation (10) gives the law of variation of dynamical variables with time in Heisenberg's picture and is Heisenberg's form for the equations of motion. It is comparable with the classical equations of motion, since these are also concerned with the variation of dynamical variables and not, like Schrödinger's form for the quantum equations of motion, with the variation of states. The classical equations of motion are

$$\frac{dq_r}{dt} = \frac{\partial H}{\partial p_r}, \quad \frac{dp_r}{dt} = -\frac{\partial H}{\partial q_r}, \quad (11)$$

$H$  being the classical Hamiltonian and the  $q$ 's and  $p$ 's a set of canonical coordinates and momenta. They give, for  $\xi$  any function of the  $q$ 's and  $p$ 's that does not contain the time  $t$  explicitly,

$$\begin{aligned} \frac{d\xi}{dt} &= \sum_r \left\{ \frac{\partial \xi}{\partial q_r} \frac{dq_r}{dt} + \frac{\partial \xi}{\partial p_r} \frac{dp_r}{dt} \right\} \\ &= \sum_r \left\{ \frac{\partial \xi}{\partial q_r} \frac{\partial H}{\partial p_r} - \frac{\partial \xi}{\partial p_r} \frac{\partial H}{\partial q_r} \right\} \\ &= [\xi, H], \end{aligned} \quad (12)$$

with the classical definition of a P.B., equation (1) of Chapter V. But equation (10) takes precisely the form (12) with the quantum definition of a P.B., equation (7) of Chapter V. We thus get an analogy between the classical and quantum equations of motion, on the basis of the analogy between classical and quantum P.B.'s, discussed in Chapter V, and we also get a justification for calling the linear operator  $H$  introduced by equation (1) the Hamiltonian of the quantum-mechanical system.

Our general derivation of equation (10) shows that the equations of motion of any dynamical system in quantum mechanics are determined by a Hamiltonian, whether the system is one that has a classical



analogue and is describable in terms of canonical coordinates and momenta or not. A system is defined mathematically by its Hamiltonian being given. When the system does have a classical analogue, it is usually permissible to assume that the Hamiltonian is the same function of the dynamical variables as in the analogous classical system.† There would be a difficulty in this, of course, if the classical Hamiltonian involved a product of factors whose quantum analogues do not commute, as one would not know in which order to put these factors in the quantum Hamiltonian, but this does not happen for most of the elementary dynamical systems whose study is important for atomic physics. In consequence we are able also largely to use the same language for describing dynamical systems in the quantum theory as in the classical theory (e.g. to talk about particles with given masses moving through given fields of force), and when given a system in classical mechanics, can usually give a meaning to 'the same' system in quantum mechanics.

A system in quantum mechanics is usually defined by its Hamiltonian being given as an algebraic function of dynamical variables, the nature of these dynamical variables being defined by their quantum conditions. This does not include the most general systems, however. It is possible to have a system whose Hamiltonian is not expressible algebraically in terms of dynamical variables, but can be specified only through its representative in some representation being given. An example of such a system is provided by the interaction of a photon with an atom, as will be dealt with in Chapter XI.

The equation of motion (12) must be generalized when  $\xi$  involves the time  $t$  explicitly as well as being a function of the dynamical variables. The generalization is, of course,

$$\frac{d\xi}{dt} = \frac{\partial \xi}{\partial t} + [\xi, H],$$

in the quantum theory as well as in the classical theory. The generalization of (10) is thus

$$i\hbar \frac{d\xi}{dt} = i\hbar \frac{\partial \xi}{\partial t} + \xi H - H\xi. \quad (13)$$

A function of the dynamical variables not involving the time explicitly is, according to (10), a constant if it commutes with  $H$ .

† This assumption is found in practice to be successful only when applied with the dynamical coordinates and momenta referring to a Cartesian system of axes and not to more general curvilinear coordinates.



It is then called a *constant of the motion*. It must commute with  $H$  at all times, which is possible usually only if  $H$  is a constant. The constancy of  $H$  in our present Heisenberg picture requires, according to (13) applied with  $\xi = H$ , that  $\partial H/\partial t = 0$ , or that  $H$  is a function† of the dynamical variables not involving the time explicitly, and therefore is a constant also in the Schrödinger picture. The result that  $H$  is a constant of the motion if  $\partial H/\partial t = 0$  is a formal expression of the law of the conservation of energy for a system in which there are no external forces. The corresponding formal expression of conservation of momentum follows from the requirement that the Hamiltonian of a system with no external forces must be an observable that is unchanged by a displacement of the type considered in § 29 and must therefore, according to equation (57) of that section with  $\xi = H$ , commute with the displacement operator, i.e. according to (59) of that section, with the total momentum. Conservation of angular momentum may be deduced in a similar way, for a system whose Hamiltonian is spherically symmetrical, with the help of the rotation operators of § 29.

We can conveniently work with a fixed representation in the Heisenberg picture only for dynamical systems whose Hamiltonian is constant. We then take the Hamiltonian itself to be diagonal. A representation of this type we call a *Heisenberg representation*, as it was introduced by *Heisenberg* in 1925. It was historically the first form of quantum mechanics to be discovered. In a Heisenberg representation every diagonal matrix represents a function of the dynamical variables that commutes with the Hamiltonian and is therefore a constant of the motion. The problem of setting up a Heisenberg representation thus reduces to the problem of finding a complete set of commuting observables, each of which is a constant of the motion, and then making these observables diagonal. The Hamiltonian itself may be one of these observables. Each of the basic states of the representation is an eigenstate of  $H$  and is therefore, according to a result of the preceding section, a stationary state.

Take a Heisenberg representation with the complete set of commuting observables  $\alpha$ , each of which is a constant of the motion, diagonal. From a theorem on page 59, the Hamiltonian  $H$ , being diagonal, must be a function of the  $\alpha$ 's, say  $H(\alpha)$ . Thus, taking

† In a generalized sense, not necessarily an algebraic function.



for definiteness the case of discrete eigenvalues for the  $\alpha$ 's, we shall have for the representative of  $H$ , from formula (30) of page 59,

$$(\alpha'|H|\alpha'') = H'\delta_{\alpha'\alpha''}, \quad (14)$$

where  $H'$  is short for  $H(\alpha')$ . If now  $\xi$  denotes any dynamical variable, or any function of the dynamical variables not involving the time explicitly, we obtain, expressing (10) in terms of representatives,

$$i\hbar \left( \alpha' \left| \frac{d\xi}{dt} \right| \alpha'' \right) = (\alpha'|\xi|\alpha'')H'' - H'(\alpha'|\xi|\alpha'')$$

or 
$$i\hbar \frac{d}{dt} (\alpha'|\xi|\alpha'') = -(H' - H'')(\alpha'|\xi|\alpha'').$$

Hence 
$$(\alpha'|\xi|\alpha'') = (\alpha'|\xi|\alpha'')_0 e^{i(H' - H'')t/\hbar}, \quad (15)$$

where  $(\alpha'|\xi|\alpha'')_0$  is independent of  $t$ . Formula (15) shows how the matrix elements representing any dynamical variable in a Heisenberg representation vary with the time. The variation is simply periodic with the frequency

$$|H' - H''|/2\pi\hbar = |H' - H''|/\hbar, \quad (16)$$

depending only on the energy difference of the two stationary states to which the matrix element refers. This result contains the essence of the Combination Law of Spectroscopy and of Bohr's Frequency Condition, according to which (16) is the frequency of the electromagnetic radiation emitted or absorbed when the system makes a transition under the influence of the radiation between the stationary states  $\alpha'$  and  $\alpha''$ , the eigenvalues of  $H$  being Bohr's energy levels. These matters will be dealt with in § 48.

The above representation with the constants of the motion  $\alpha$  diagonal is fixed in the Heisenberg picture, and is thus moving in the Schrödinger picture. We could introduce a representation with the  $\alpha$ 's diagonal, which is fixed in the Schrödinger picture and is thus moving in the Heisenberg picture. The two representations would differ only in the phase factors. The representative,  $(\alpha'|\xi|\alpha'')^*$  say, of a dynamical variable  $\xi$  in the latter representation would not vary with the time and would thus, according to (15), be connected with the representative  $(\alpha'|\xi|\alpha'')$  in the former representation by the law

$$(\alpha'|\xi|\alpha'')^* = (\alpha'|\xi|\alpha'')e^{-i(H' - H'')t/\hbar},$$

with neglect of a possible constant phase factor. Hence the repre-



sentative  $(\alpha' |)^*$  of a  $\psi$  in the latter representation would be connected with its representative  $(\alpha' |)$  in the former by

$$(\alpha' |)^* = (\alpha' |) e^{-iH' t/\hbar}. \quad (17)$$

### 33. The Action Principle

The analogy between Heisenberg's form for the equations of motion (10) and the classical equation of motion (12) enables us to pursue the analogy between classical dynamics and quantum dynamics further and to see how all the main principles and results of the classical theory reappear in the quantum theory in a generalized form.

If we denote by  $\xi_t$  the dynamical variable  $\xi$  at time  $t$ , then equation (10) gives us, for  $\delta t$  infinitesimal,

$$i\hbar(\xi_{t+\delta t} - \xi_t) = \delta t(\xi H - H\xi),$$

or

$$\xi_{t+\delta t} - \xi_t = i \delta t/\hbar \cdot (H\xi - \xi H).$$

Comparing this with (76) of Chapter V, we see that *the dynamical variables at time  $t + \delta t$  are connected with the dynamical variables at time  $t$  by an infinitesimal contact transformation*. Thus the changing of the dynamical variables under the equation of motion (10) may be regarded as the continual development of a contact transformation. After the lapse of a finite time the dynamical variables will be connected with the initial dynamical variables by a finite contact transformation. These results are formally the same as in classical mechanics. One might expect them in quantum mechanics simply from the requirement that the quantum conditions must hold throughout all time, the only general transformations which leave invariant quantum conditions, or any algebraic equations, being the contact transformations of § 30.

If the Hamiltonian is a constant, the contact transformation connecting the dynamical variables at time  $t$ ,  $\xi_t$ , with the initial dynamical variables  $\xi_0$ , may be written

$$\xi_t = e^{iHt/\hbar} \xi_0 e^{-iHt/\hbar}. \quad (18)$$

To verify this equation, we note that it obviously holds for  $t = 0$ , and when differentiated with respect to  $t$  gives

$$\frac{d\xi_t}{dt} = \frac{de^{iHt/\hbar}}{dt} \xi_0 e^{-iHt/\hbar} + e^{iHt/\hbar} \xi_0 \frac{de^{-iHt/\hbar}}{dt},$$

or

$$\begin{aligned} i\hbar \frac{d\xi_t}{dt} &= -He^{iHt/\hbar} \xi_0 e^{-iHt/\hbar} + e^{iHt/\hbar} \xi_0 e^{-iHt/\hbar} H \\ &= -H\xi_t + \xi_t H, \end{aligned}$$



which is just the equation of motion (10). Equation (18) thus provides an explicit solution in symbolic form of the differential equation (10). This solution, like equation (5), is not often useful in practice, owing to the difficulty of evaluating the exponentials.

In the Heisenberg picture in which the states are represented by fixed  $\psi$ -vectors and the dynamical variables by varying linear operators, we may introduce a fixed representation in which the diagonal observables are dynamical variables at some definite time  $t$ . They may, for instance, be the coordinates at time  $t$ ,  $q_t$  say, assuming the system to have canonical coordinates and momenta. We should then have one representation for each time  $t$  and should have a transformation function  $(q'_t|q'_T)$  connecting the representations referring to two different times  $t$  and  $T$ . The law of transformation for the representative of a  $\psi$ -symbol will be

$$(q'_t|) = \int (q'_t|q'_T) dq'_T (q'_T|).$$

If in this equation we vary  $t$  keeping  $T$  and the function  $(q'_T|)$  fixed, the resulting  $(q'_t|)$  will give us the representative at various times of a fixed  $\psi$  referred to the moving axes of the Heisenberg picture. This must be the same as the representative of a moving  $\psi$ , representing a state as it varies with time, referred to the fixed axes of the Schrödinger picture, and must therefore satisfy Schrödinger's wave equation (3), i.e.

$$i\hbar \frac{d}{dt} \int (q'_t|q'_T) dq'_T (q'_T|) = \int \int (q'_t|H|q''_t) dq''_t (q'_t|q'_T) dq'_T (q'_T|).$$

This holds for an arbitrary function  $(q'_T|)$  and hence

$$i\hbar \frac{d}{dt} (q'_t|q'_T) = \int (q'_t|H|q''_t) dq''_t (q''_t|q'_T). \quad (19)$$

Thus the transformation function  $(q'_t|q'_T)$ , considered as a function of the variables  $q'_t$ , is a solution of Schrödinger's wave equation. Similarly, considered as a function of the variables  $q'_T$ , it satisfies an equation of the form (4), namely

$$-i\hbar \frac{d}{dT} (q'_t|q'_T) = \int (q'_t|q''_T) dq''_T (q''_T|H|q'_T). \quad (20)$$

From the analogy between classical and quantum contact transformations discussed in § 30, we see that  $(q'_t|q'_T)$  corresponds in the classical theory to  $e^{iS/\hbar}$ , where  $S$  is Hamilton's principal function for



the time interval  $T$  to  $t$ , equal to the time-integral of the Lagrangian  $L$ ,

$$S = \int_T^t L dt. \quad (21)$$

Taking an infinitesimal time interval  $t$  to  $t + \delta t$ , we see that  $(q'_{t+\delta t} | q'_t)$  corresponds to  $e^{iL\delta t/\hbar}$ . This result gives probably the most fundamental quantum analogue for the classical Lagrangian function. It is preferable for the sake of the analogy to consider the classical Lagrangian as a function of the coordinates at time  $t$  and the coordinates at time  $t + \delta t$ , instead of a function of the coordinates and velocities at time  $t$ .

There is an important action principle in classical mechanics concerning Hamilton's principal function (21). It says that this function remains stationary for small variations of the trajectory of the system which do not alter the end points, i.e. for small variations of the  $q$ 's at all intermediate times between  $T$  and  $t$  with  $q_T$  and  $q_t$  fixed. Let us see what it corresponds to in the quantum theory.

Put

$$\exp\left\{i \int_{t_a}^{t_b} L dt/\hbar\right\} = \exp\{iS(t_b, t_a)/\hbar\} = B(t_b, t_a), \quad (22)$$

so that  $B(t_b, t_a)$  corresponds to  $(q'_{t_b} | q'_{t_a})$  in the quantum theory. Now suppose the time interval  $T \rightarrow t$  to be divided up into a large number of small time intervals  $T \rightarrow t_1$ ,  $t_1 \rightarrow t_2, \dots$ ,  $t_{m-1} \rightarrow t_m$ ,  $t_m \rightarrow t$ , by the introduction of a sequence of intermediate times  $t_1, t_2, \dots, t_m$ . Then

$$B(t, T) = B(t, t_m)B(t_m, t_{m-1}) \dots B(t_2, t_1)B(t_1, T). \quad (23)$$

The corresponding quantum equation, which follows from the composition law (43) of Chapter III, is

$$(q'_t | q'_T) = \int \dots \int (q'_t | q'_m) dq'_m (q'_m | q'_{m-1}) dq'_{m-1} \dots (q'_2 | q'_1) dq'_1 (q'_1 | q'_T), \quad (24)$$

$q'_k$  being written for  $q'_{t_k}$  for brevity. At first sight there does not seem to be any close correspondence between (23) and (24). We must, however, analyse the meaning of (23) rather more carefully. We must regard each factor  $B$  as a function of the  $q$ 's at the two ends of the time interval to which it refers. This makes the right-hand side of (23) a function, not only of  $q_t$  and  $q_T$ , but also of all the intermediate  $q$ 's. Equation (23) is valid only when we substitute for the intermediate  $q$ 's in its right-hand side their values for the real trajectory, small variations in which values leave  $S(t, T)$  stationary and therefore also, from (22), leave  $B(t, T)$  stationary. It is the process of



substituting these values for the intermediate  $q$ 's which corresponds to the integrations over all values for the intermediate  $q$ 's in (24). The quantum analogue of Hamilton's action principle is thus absorbed in the composition law (24) and the classical requirement that the values of the intermediate  $q$ 's shall make  $S(t, T)$  stationary corresponds to the condition in quantum mechanics that all values of the intermediate  $q$ 's are important in proportion to their contribution to the integral in (24).

Let us see how (23) can be a limiting case of (24) for  $\hbar$  small. We must suppose the integrand in (24) to be of the form  $e^{iF/\hbar}$ , where  $F$  is a function of  $q'_T, q'_1, q'_2, \dots, q'_m, q'_t$  which remains continuous as  $\hbar$  tends to zero, so that the integrand is a rapidly oscillating function when  $\hbar$  is small. The integral of such a rapidly oscillating function will be extremely small, except for the contribution arising from a region in the domain of integration where comparatively large variations in the  $q'_k$  produce only very small variations in  $F$ . Such a region must be the neighbourhood of a point where  $F$  is stationary for small variations of the  $q'_k$ . Thus the integral in (24) is determined essentially by the value of the integrand at a point where the integrand is stationary for small variations of the intermediate  $q$ 's, and so (24) goes over into (23).

### 34. The Motion of Wave Packets

The comparison between classical and quantum mechanics may be discussed with reference to a wave function,  $(q'|)$  or  $(q'_t|)$ , instead of, as we did above, with reference to a transformation function  $(q'_t|q'_T)$ . The transformation function  $(q'_t|q'_T)$  is like a wave function in its dependence on the variables  $q'_t$ , as is shown by equation (19), and if we are interested only in the variables  $q'_t$  and not in  $q'_T$ , the natural thing to do is to study a wave function instead of the transformation function. The resulting simplification will enable us to push the comparison to a higher degree of accuracy without getting laborious calculations.

Let us take a quantum dynamical system having a classical analogue and therefore describable with canonical coordinates and momenta and assume that its Hamiltonian is a function of the coordinates and momenta expressible as a power series in the momenta. The Hamiltonian is thus expressible as a sum of terms, each of which is a product of various powers of the momenta and of



various functions of the coordinates, with no restriction on the order of the factors. To facilitate comparison with the classical theory we shall suppose that these functions of the coordinates are all real and that the Hamiltonian does not involve  $i$  in any way. This condition does not mean any loss of generality in our dynamical system, since if it does not hold we can make it hold by simplifying the expression for the Hamiltonian in the following way. We can certainly express the Hamiltonian as

$$H = H_1 + iH_2, \quad (25)$$

in which  $H_1$  and  $H_2$  involve the coordinates only through real functions and do not involve  $i$ .  $H_1$  and  $iH_2$  individually need not be Hermitian, although, of course,  $H$  must be. Thus, taking the conjugate complex of equation (25), we get

$$H = \bar{H}_1 - i\bar{H}_2. \quad (26)$$

According to the rules of § 15 for obtaining conjugate complexes,  $\bar{H}_1$  and  $\bar{H}_2$  will be just  $H_1$  and  $H_2$  with the factors in all their terms in the reverse order, since each factor by itself is Hermitian. From (25) and (26) we have

$$H = \frac{1}{2}(H_1 + \bar{H}_1) + \frac{1}{2}i(H_2 - \bar{H}_2). \quad (27)$$

For each term in  $H_2$  there will be a corresponding term in  $\bar{H}_2$  consisting of the same factors in the reverse order and the difference of two such terms can be reduced, by means of general theorems on P.B.'s given in § 25 and of equation (38) in § 27, to  $i\hbar$  times an expression not involving  $i$  in any way. By carrying out this reduction for all the terms in  $H_2 - \bar{H}_2$  in (27), we get  $H$  in the required form not involving  $i$ . It should be noted that in this form  $H$  remains unchanged if we reverse the order of the factors in every term.

Since  $H$  is expressible as a power series in the momenta, in a representation in which the coordinates  $q_r$  are diagonal it will be represented by a differential operator of the form (36) of Chapter V, and thus Schrödinger's wave equation (3) will read

$$i\hbar \frac{d}{dt}(q'|) = H\left(q', -i\hbar \frac{\partial}{\partial q'}\right)(q'|). \quad (28)$$

Let us study the nature of the solution of (28) in the limiting case of  $\hbar$  very small. We try to get a solution in the form of waves

$$(q'|) = e^{iS/\hbar} A, \quad (29)$$

where  $S$  and  $A$  are real functions of the  $q$ 's and  $t$  which give the



phase and amplitude respectively. The appearance of the factor  $A$  here marks a step towards higher accuracy than we had in the preceding section.

With (29), the effect of the operator  $-i\hbar \partial/\partial q'_r$  on the wave function  $(q'|)$  is

$$-i\hbar \frac{\partial}{\partial q'_r}(q'|) = e^{iS/\hbar} \left( \frac{\partial S}{\partial q'_r} - i\hbar \frac{\partial}{\partial q'_r} \right) A \quad (30)$$

and that of the operator  $i\hbar d/dt$  is

$$i\hbar \frac{d}{dt}(q'|) = e^{iS/\hbar} \left( -\frac{\partial S}{\partial t} + i\hbar \frac{\partial}{\partial t} \right) A.$$

If  $f$  is any function of the operators  $-i\hbar \partial/\partial q'_r$  expressible as a power series, we find readily by repeated applications of (30)

$$f\left(-i\hbar \frac{\partial}{\partial q'_r}\right)(q'|) = e^{iS/\hbar} f\left(\frac{\partial S}{\partial q'_r} - i\hbar \frac{\partial}{\partial q'_r}\right) A.$$

Thus when we substitute the expression (29) for  $(q'|)$  into (28) we shall get, after cancelling the factor  $e^{iS/\hbar}$ ,

$$\left(-\frac{\partial S}{\partial t} + i\hbar \frac{\partial}{\partial t}\right) A = H\left(q', \frac{\partial S}{\partial q'} - i\hbar \frac{\partial}{\partial q'}\right) A. \quad (31)$$

The operator on the right-hand side here is a power series in the  $(\partial S/\partial q' - i\hbar \partial/\partial q')$ 's and is thus a power series in the  $(i\hbar \partial/\partial q')$ 's. We shall now neglect  $\hbar^2$  and thus neglect terms of higher degree than the first in this power series in the  $(i\hbar \partial/\partial q')$ 's. The terms of zero degree and of the first degree are real and pure imaginary respectively, and therefore we shall have to equate the results of their operating on  $A$  to the real and pure imaginary parts of the left-hand side of (31) respectively.

Equating the real parts on both sides of (31) we get, after cancelling the factor  $A$ ,

$$-\frac{\partial S}{\partial t} = H\left(q', \frac{\partial S}{\partial q'}\right). \quad (32)$$

This is just the Hamilton-Jacobi equation of classical mechanics, with  $S$  as Hamilton's principal function, and is what we should expect from our work in the preceding section.

Let us now pick out the terms of first degree in  $i\hbar \partial/\partial q'_r$  in the operator on the right-hand side of (31). These terms will give us an operator of the general form

$$\sum_k X_k i\hbar \frac{\partial}{\partial q'_r} Y_k, \quad (33)$$



where the  $X$ 's and  $Y$ 's are functions of the  $q$ 's. The total coefficient of  $i\hbar \partial/\partial q'_r$  in (33), namely  $\sum_k X_k Y_k$ , must be equal to

$$\sum_k X_k Y_k = -\frac{\partial H(q', \partial S/\partial q')}{\partial(\partial S/\partial q'_r)}, \quad (34)$$

but we cannot immediately use this result on account of the sandwiched positions of  $i\hbar \partial/\partial q'_r$  in (33). We must first use the condition mentioned above, that the expression for the Hamiltonian in co-ordinates and momenta remains unchanged if we reverse the order of the factors in every term. This means that the operator on the right-hand side of (31), and hence also the operator (33), will remain unchanged if we reverse the order of the factors in every term. Thus

$$\begin{aligned} \sum_k X_k i\hbar \frac{\partial}{\partial q'_r} Y_k &= \sum_k Y_k i\hbar \frac{\partial}{\partial q'_r} X_k \\ &= \frac{1}{2} \sum_k \left\{ X_k i\hbar \frac{\partial}{\partial q'_r} Y_k + Y_k i\hbar \frac{\partial}{\partial q'_r} X_k \right\} \\ &= \frac{1}{2} \sum_k \left\{ X_k Y_k i\hbar \frac{\partial}{\partial q'_r} + i\hbar \frac{\partial}{\partial q'_r} Y_k X_k \right\} \\ &= -\frac{1}{2} \left\{ \frac{\partial H(q', \partial S/\partial q')}{\partial(\partial S/\partial q'_r)} i\hbar \frac{\partial}{\partial q'_r} + i\hbar \frac{\partial}{\partial q'_r} \frac{\partial H(q', \partial S/\partial q')}{\partial(\partial S/\partial q'_r)} \right\} \end{aligned} \quad (35)$$

from (34). We must now equate the result of the operator (35) summed for all values of  $r$ , operating on  $A$ , to the pure imaginary part of the left-hand side of (31). This gives

$$\frac{\partial A}{\partial t} = -\frac{1}{2} \sum_r \left\{ \frac{\partial H(q', \partial S/\partial q')}{\partial(\partial S/\partial q'_r)} \frac{\partial}{\partial q'_r} + \frac{\partial}{\partial q'_r} \frac{\partial H(q', \partial S/\partial q')}{\partial(\partial S/\partial q'_r)} \right\} A,$$

which, on multiplication by  $2A$ , reduces to

$$\frac{\partial A^2}{\partial t} = - \sum_r \frac{\partial}{\partial q'_r} \left\{ \frac{\partial H(q', \partial S/\partial q')}{\partial(\partial S/\partial q'_r)} A^2 \right\}. \quad (36)$$

This is the equation for the amplitude  $A$  of the wave function. To get an understanding of its significance, let us suppose we have a fluid moving in the space of the variables  $q'$ , the density of the fluid at any point and time being  $A^2$  and its velocity

$$\frac{dq_r}{dt} = \frac{\partial H(q', \partial S/\partial q')}{\partial(\partial S/\partial q'_r)}. \quad (37)$$

Equation (36) is then just the equation of conservation for such a fluid. There is one velocity function (37) for each function  $S$  satisfying (32).



Let us take a solution of (36) for which at some definite time the density  $A^2$  vanishes everywhere outside a certain small region. We may suppose this region to move with the fluid, its velocity at each point being given by (37), and then the equation of conservation (36) will require the density always to vanish outside the region. There is a limit to how small the region may be, imposed by the approximation we made above in neglecting  $\hbar^2$  in the operator in the right-hand side of (31). This approximation is valid only provided

$$\hbar \frac{\partial}{\partial q'} A \ll \frac{\partial S}{\partial q'} A,$$

which requires that  $A$  shall vary by an appreciable fraction of itself only through a range of  $q'$  in which  $S$  varies by many times  $\hbar$ , i.e. a range consisting of many wave-lengths of the wave function (29). Our solution is then a wave packet of the type discussed in § 28 and remains so for all time.

We thus get a wave function representing a state† for which the coordinates and momenta have approximate numerical values throughout all time. Such a state in quantum theory corresponds to the states with which classical theory deals. The motion of our wave packet is given by equation (37) and is therefore, from the Hamilton-Jacobi theory of classical mechanics in which the momenta  $p_r$  are replaced by  $\partial S / \partial q_r$ , just along the classical trajectory. This gives us a justification, of a less formal type than the analogy discussed in § 32, for considering the classical equations of motion as the limiting form of the quantum ones when  $\hbar \rightarrow 0$ .

By a more accurate solution of the wave equation one can show that the accuracy with which the coordinates and momenta simultaneously have numerical values cannot remain permanently as favourable as the limit allowed by Heisenberg's principle of uncertainty, equation (42) of Chapter V, but if it is initially so it will become less favourable, the wave packet undergoing a spreading.‡

### 35. The Free Particle

The most fundamental and elementary application of quantum mechanics is to the system consisting merely of a free particle, or

† The word 'state' is here used with its space-time meaning.

‡ See Kennard, *Z. f. Physik*, **44** (1927), 344; Darwin, *Proc. Roy. Soc. A*, **117** (1927), 258.



particle not acted on by any forces. The problem is still very simple when we take into account, as we shall do here, the relativistic variation of the mass of the particle with its velocity. We shall use as dynamical variables the three Cartesian coordinates of the particle  $x, y, z$ , and their conjugate momenta  $p_x, p_y, p_z$ . In terms of these variables, the Hamiltonian in classical mechanics, equal to the energy, is

$$H = c(m^2c^2 + p_x^2 + p_y^2 + p_z^2)^{\frac{1}{2}}, \quad (38)$$

where  $m$  is the rest-mass of the particle and  $c$  is the velocity of light. We assume the Hamiltonian to be of the same form in quantum mechanics, the square root now being interpreted as the positive square root defined at the end of § 11.

From the quantum conditions (9) of Chapter V,  $p_x$  commutes with  $p_y$  and  $p_z$ , and hence, from the theorem given at the end of § 16,  $p_x$  commutes with any function of  $p_x, p_y$ , and  $p_z$  and therefore with  $H$ . It follows that  $p_x$  is a constant of the motion. Similarly  $p_y$  and  $p_z$  are constants of the motion. These results are the same as in the classical theory. Again, the equation of motion for a coordinate,  $x$  say, is, according to (10),

$$i\hbar \dot{x} = i\hbar \frac{dx}{dt} = xc(m^2c^2 + p_x^2 + p_y^2 + p_z^2)^{\frac{1}{2}} - c(m^2c^2 + p_x^2 + p_y^2 + p_z^2)^{\frac{1}{2}}x.$$

The right-hand side here can be evaluated by means of formula (38) of Chapter V with the roles of coordinates and momenta interchanged, so that it reads

$$fq_s - q_sf = -i\hbar \frac{\partial f}{\partial p_s}, \quad (39)$$

$f$  now being any function of the  $p$ 's. This gives

$$\dot{x} = \frac{\partial}{\partial p_x} c(m^2c^2 + p_x^2 + p_y^2 + p_z^2)^{\frac{1}{2}} = \frac{c^2 p_x}{H}.$$

Similarly, 
$$\dot{y} = \frac{c^2 p_y}{H} \quad \dot{z} = \frac{c^2 p_z}{H}. \quad (40)$$

These equations of motion are of the same form as in the classical theory.

Let us consider a state that is an eigenstate of the momenta, belonging to the eigenvalues  $p'_x, p'_y, p'_z$ . This state must be an eigenstate of the Hamiltonian, belonging to the eigenvalue

$$H' = c(m^2c^2 + p'^2_x + p'^2_y + p'^2_z)^{\frac{1}{2}} \quad (41)$$

and must therefore be a stationary state. The possible values for  $H'$



are all numbers from  $mc^2$  to  $\infty$ , as in the classical theory. In a representation with the coordinates  $x, y, z$  diagonal, the representative of our stationary state at any time  $t$  will be, from (39) of Chapter V, of the form

$$(x'y'z'|) = ae^{i(x'p'_x+y'p'_y+z'p'_z)/\hbar},$$

where  $a$  is independent of  $x', y', z'$  but may depend on the time  $t$ . From (7) we see that  $a$  varies with  $t$  according to the simple harmonic law

$$a = a_0 e^{-iHt/\hbar},$$

where  $a_0$  is a constant, and hence

$$(x'y'z'|) = a_0 e^{i(p'_x x' + p'_y y' + p'_z z' - Ht)/\hbar}. \quad (42)$$

Formula (42) gives the wave function representing a state with definite momentum, for the problem of a free particle. It could of course have been obtained alternatively from a direct solution of Schrödinger's wave equation (3). It is of the form of plane waves in space-time. The frequency of the waves is

$$\nu = H'/\hbar, \quad (43)$$

their wave-length is

$$\lambda = \hbar/(p'^2_x + p'^2_y + p'^2_z)^{\frac{1}{2}} = \hbar/P', \quad (44)$$

$P'$  being the length of the vector  $(p'_x, p'_y, p'_z)$ , and their motion is in the direction specified by the vector  $(p'_x, p'_y, p'_z)$  with the velocity

$$u = \lambda\nu = H'/P' = c^2/v, \quad (45)$$

$v$  being the classical velocity of the particle corresponding to the momentum  $(p'_x, p'_y, p'_z)$ . Equations (43), (44), and (45) are easily seen to hold in all Lorentz frames of reference, the expression on the right-hand side of (42) being, in fact, relativistically invariant with  $p'_x, p'_y, p'_z$  and  $H'$  as the components of a 4-vector. These properties of relativistic invariance led *de Broglie*, before the discovery of quantum mechanics, to postulate the existence of waves such as (42) associated with the motion of any particle. They are therefore known as *de Broglie waves*. In the limiting case when the rest-mass  $m$  is made to tend to zero, the classical velocity of the particle  $v$  becomes equal to  $c$  and hence, from (45), the wave velocity also becomes  $c$ . The waves then become identical with the light-waves associated with a photon, except for the fact that they contain no reference to the polarization and involve a complex exponential instead of sines and cosines. Formulas (43) and (44) are still valid, connecting the frequency of the light-waves with the energy of the photon and the wave-length of the light-waves with the momentum of the photon.



For the state represented by (42), the probability of the particle being found in any specified small volume when an observation of its position is made is proportional to  $|(x'y'z')|^2$  and is thus independent of the position of the volume. This provides an example of Heisenberg's principle of uncertainty, the state being one for which the momentum is accurately given and for which, in consequence, the position is completely unknown. Such a state is, of course, a limiting case which never occurs in practice. The states usually met with in practice are those represented by wave packets, which may be formed by superposing a number of waves of the type (42) belonging to slightly different values of  $(p'_x, p'_y, p'_z)$ . The ordinary formula in hydrodynamics for the velocity of such a wave packet, i.e. the *group velocity* of the waves, is

$$\frac{dv}{d(1/\lambda)}$$

which gives, from (43) and (44)

$$\frac{dH'}{dP'} = c \frac{d}{dP'} (m^2 c^2 + P'^2)^{\frac{1}{2}} = \frac{c^2 P'}{H'} = v. \quad (46)$$

This is just the classical velocity of the particle and confirms the general theory of the preceding section.

### 36. The Harmonic Oscillator

As another example of a simple system treated according to quantum mechanics, we may take the harmonic oscillator, neglecting relativistic variation of mass with velocity. We have as variables only one coordinate  $q$  and its conjugate momentum  $p$  and we take the Hamiltonian to be, as in the classical theory,

$$H = \frac{1}{2m} (p^2 + m^2 \omega^2 q^2), \quad (47)$$

$m$  being the mass of the oscillating particle and  $\omega$  being  $2\pi$  times the frequency. With this Hamiltonian it is easily verified that the equations of motion for  $q$  and  $p$  are

$$\dot{q} = p/m, \quad \dot{p} = -m\omega^2 q, \quad (48)$$

precisely as in the classical theory.

We must now determine the eigenvalues of the Hamiltonian. This could be done directly by solving the differential equation (37) of Chapter V. An alternative method, based on more primitive arguments, is as follows. We have from straightforward non-commutative algebra, with the help of the quantum condition (12) of Chapter V,



$$\begin{aligned}
 (p+im\omega q)(p-im\omega q) &= p^2+m^2\omega^2q^2+im\omega(qp-pq) \\
 &= p^2+m^2\omega^2q^2-m\hbar\omega \\
 &= 2mH-m\hbar\omega,
 \end{aligned} \tag{49}$$

and similarly,

$$(p-im\omega q)(p+im\omega q) = 2mH+m\hbar\omega. \tag{50}$$

Hence

$$\begin{aligned}
 (2mH-m\hbar\omega)(p+im\omega q) &= (p+im\omega q)(p-im\omega q)(p+im\omega q) \\
 &= (p+im\omega q)(2mH+m\hbar\omega).
 \end{aligned} \tag{51}$$

We now introduce a Heisenberg representation in which  $H$  is diagonal. We shall assume that  $H$  by itself forms a complete set of commuting observables and its eigenvalues can therefore be used for labelling coordinates in the representation. The justification for this assumption is that it leads, as we shall see, without inconsistency to definite representatives for  $q$  and  $p$ . Expressing (51) in terms of representatives, we obtain

$$\begin{aligned}
 \{2mH'-m\hbar\omega\}(H'|p+im\omega q|H'') &= (H'|p+im\omega q|H'')\{2mH''+m\hbar\omega\} \\
 \text{or} \quad \{H'-H''-\hbar\omega\}(H'|p+im\omega q|H'') &= 0.
 \end{aligned} \tag{52}$$

This shows that all the matrix elements  $(H'|p+im\omega q|H'')$  of the representative of  $p+im\omega q$  vanish except those for which

$$H'-H''-\hbar\omega = 0. \tag{53}$$

Taking the conjugate complex of this result in accordance with (18) of Chapter III, we see that all the matrix elements  $(H''|p-im\omega q|H')$  of the representative of  $p-im\omega q$  vanish except those for which (53) holds. It follows that in the equation

$$\begin{aligned}
 \sum_{H''} (H'|p+im\omega q|H'')(H''|p-im\omega q|H') &= (H'|2mH-m\hbar\omega|H') \\
 &= 2mH'-m\hbar\omega = 2m\{H'-\tfrac{1}{2}\hbar\omega\}
 \end{aligned} \tag{54}$$

which we obtain by expressing (49) in terms of representatives and taking a diagonal matrix element of each side, referring to an arbitrarily chosen eigenvalue  $H'$ , all the terms in the sum on the left-hand side vanish except (at most) the one for which  $H'' = H' - \hbar\omega$ , if  $H' - \hbar\omega$  is an eigenvalue of  $H$ , and if it is not, then every term on the left-hand side of (54) vanishes without exception. In the first case  $H' - \frac{1}{2}\hbar\omega$  is positive or zero, since  $(H'|p+im\omega q|H' - \hbar\omega)$  and  $(H' - \hbar\omega|p-im\omega q|H')$  are conjugate complex numbers, and in the second  $H' - \frac{1}{2}\hbar\omega$  is certainly zero. We can therefore draw the conclusions that, if  $H'$  is any eigenvalue of  $H$ , then  $H'$  is positive and either



$H' - \hbar\omega$  is another eigenvalue or  $H' = \frac{1}{2}\hbar\omega$ . Similarly, by expressing (50) in terms of representatives and taking the diagonal matrix element of each side referring to  $H'$ , we can draw the conclusion that either  $H' + \hbar\omega$  is another eigenvalue or  $H' = -\frac{1}{2}\hbar\omega$ . The second alternative here is ruled out, since  $H'$  must always be positive. It follows finally from all this that the only possible set of eigenvalues for  $H$  is the series

$$\frac{1}{2}\hbar\omega, \frac{3}{2}\hbar\omega, \frac{5}{2}\hbar\omega, \frac{7}{2}\hbar\omega, \dots, \quad (55)$$

extending to infinity. These are the energy levels for the simple harmonic oscillator.

We can now easily obtain the representatives of  $q$  and  $p$ . Equation (54) reduces to

$$(H'|p + im\omega q|H' - \hbar\omega)(H' - \hbar\omega|p - im\omega q|H') = 2m\{H' - \frac{1}{2}\hbar\omega\}.$$

The two factors on the left here are conjugate complex numbers and hence

$$\begin{aligned} (H'|p + im\omega q|H' - \hbar\omega) &= (2m)^{\frac{1}{2}}\{H' - \frac{1}{2}\hbar\omega\}^{\frac{1}{2}}e^{i\gamma'} \\ (H' - \hbar\omega|p - im\omega q|H') &= (2m)^{\frac{1}{2}}\{H' - \frac{1}{2}\hbar\omega\}^{\frac{1}{2}}e^{-i\gamma'}, \end{aligned}$$

where  $\gamma'$  is some real number, which may be a function of  $H'$ . From (15) we see that  $(H'|p + im\omega q|H' - \hbar\omega)$  must vary with  $t$  according to the law

$$(H'|p + im\omega q|H' - \hbar\omega) = \text{const. } e^{i\omega t},$$

and hence  $\gamma'$  must vary with  $t$  according to the law

$$\gamma' = \omega t + \gamma'_0,$$

where  $\gamma'_0$  is a constant. We can make  $\gamma'_0$  zero by a suitable choice of the phase factors of our representation. We then have

$$\left. \begin{aligned} (H'|p + im\omega q|H' - \hbar\omega) &= (2m)^{\frac{1}{2}}\{H' - \frac{1}{2}\hbar\omega\}^{\frac{1}{2}}e^{i\omega t} \\ (H' - \hbar\omega|p - im\omega q|H') &= (2m)^{\frac{1}{2}}\{H' - \frac{1}{2}\hbar\omega\}^{\frac{1}{2}}e^{-i\omega t}. \end{aligned} \right\} \quad (56)$$

These formulas give all the non-vanishing matrix elements of the representatives of  $p + im\omega q$  and  $p - im\omega q$ , and thus of the representatives of  $p$  and  $q$ .

In the classical treatment of periodic and multiply-periodic dynamical systems it is often convenient to make use of action and angle variables. We can introduce corresponding variables in the quantum theory. In our present problem of the harmonic oscillator we can define the action variable  $J$  by

$$J = H/\omega - \frac{1}{2}\hbar. \quad (57)$$

It is a constant of the motion and its eigenvalues are integral multiples



of  $\hbar$  greater than or equal to zero. Thus its matrix representative in the Heisenberg representation is

$$\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & . & . \\ 0 & \hbar & 0 & 0 & 0 & . & . \\ 0 & 0 & 2\hbar & 0 & 0 & . & . \\ 0 & 0 & 0 & 3\hbar & 0 & . & . \\ 0 & 0 & 0 & 0 & 4\hbar & . & . \\ . & . & . & . & . & . & . \end{vmatrix}$$

when the rows and columns are arranged in order of ascending energy-values. To define the angle variable we introduce the two matrices

$$\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & . & . \\ 1 & 0 & 0 & 0 & 0 & . & . \\ 0 & 1 & 0 & 0 & 0 & . & . \\ 0 & 0 & 1 & 0 & 0 & . & . \\ 0 & 0 & 0 & 1 & 0 & . & . \\ . & . & . & . & . & . & . \end{vmatrix} \quad \begin{vmatrix} 0 & 1 & 0 & 0 & 0 & . & . \\ 0 & 0 & 1 & 0 & 0 & . & . \\ 0 & 0 & 0 & 1 & 0 & . & . \\ 0 & 0 & 0 & 0 & 1 & . & . \\ 0 & 0 & 0 & 0 & 0 & . & . \\ . & . & . & . & . & . & . \end{vmatrix}$$

in which the non-vanishing elements are just to the left and just to the right of the principal diagonal respectively, and call the variables that they represent at time  $t = 0$ ,  $e^{iw}$  and  $e^{-iw}$  respectively. These two matrices, according to § 15, represent conjugate complex dynamical variables, in agreement with what is implied by the notation of  $e^{iw}$  and  $e^{-iw}$ . This notation implies further, however, that the two matrices are the reciprocals of one another and this is not altogether true. The matrix representing the product  $e^{-iw}e^{iw}$  is, in fact, just the unit matrix, but that representing  $e^{iw}e^{-iw}$  differs from the unit matrix through having zero for its first diagonal element. Thus

$$e^{-iw}e^{iw} = 1 \quad e^{iw}e^{-iw} \neq 1. \quad (58)$$

The variables  $e^{iw}$ ,  $e^{-iw}$ , defined above through their matrix representatives, are the best quantum analogues that we can get to the exponentials of  $i$  and  $-i$  times the angle variable of the classical theory. They have many properties analogous to those of their classical counterparts, and their only defect is that  $e^{iw}e^{-iw}$  is not precisely equal to unity. Thus, for example, we obtain at once from the matrices the relations

$$\left. \begin{aligned} J e^{iw} &= e^{iw}(J + \hbar) \\ J e^{-iw} &= e^{-iw}(J - \hbar), \end{aligned} \right\} \quad (59)$$



which are equivalent to the classical relations,

$$[e^{iw}, J] = ie^{iw} \quad [e^{-iw}, J] = -ie^{-iw}.$$

Equations (59), when compared with equation (17) of Chapter V with  $c = \pm 1$ , are seen to be consistent with the view that  $J$  and  $w$  are a pair of canonically conjugate dynamical variables satisfying the quantum condition

$$wJ - Jw = i\hbar,$$

although actually this relation is meaningless since we cannot define  $w$  itself but only  $e^{\pm iw}$ . Again, the dynamical variable  $e^{iw}$  at an arbitrary time  $t$  must be represented by a matrix whose elements vary with  $t$  according to the Heisenberg law  $e^{i(H' - H'')t/\hbar}$ . Since all the matrix elements vanish except those referring to consecutive energy-levels for which  $H' - H'' = \hbar\omega$ , every matrix element will vary with the time according to the law  $e^{i\omega t}$ . This corresponds to the fact that in the classical theory  $w$  increases linearly with  $t$  at the rate  $\omega$ .

The dynamical variables  $q$  and  $p$  can be expressed in terms of the action and angle variables. From (56) we see that

$$\begin{aligned} p + im\omega q &= (2m)^{\frac{1}{2}} \{H - \tfrac{1}{2}\hbar\omega\}^{\frac{1}{2}} e^{iw} \\ &= (2m\omega)^{\frac{1}{2}} J^{\frac{1}{2}} e^{iw} \\ p - im\omega q &= (2m)^{\frac{1}{2}} e^{-iw} \{H - \tfrac{1}{2}\hbar\omega\}^{\frac{1}{2}} \\ &= (2m\omega)^{\frac{1}{2}} e^{-iw} J^{\frac{1}{2}}. \end{aligned}$$

Thus

$$\left. \begin{aligned} p &= (\tfrac{1}{2}m\omega)^{\frac{1}{2}} \{J^{\frac{1}{2}} e^{iw} + e^{-iw} J^{\frac{1}{2}}\} \\ q &= (2m\omega)^{-\frac{1}{2}} \{-iJ^{\frac{1}{2}} e^{iw} + ie^{-iw} J^{\frac{1}{2}}\}. \end{aligned} \right\} \quad (60)$$

We see from these equations that  $q$  and  $p$ , when expressed in terms of the action and angle variables, involve them only through the two combinations  $J^{\frac{1}{2}} e^{iw}$  and  $e^{-iw} J^{\frac{1}{2}}$ . Further, all dynamical variables that we ordinarily have to deal with to obtain physical results are algebraic functions of  $q$  and  $p$  and therefore, when expressed in terms of the action and angle variables, will involve them only through the two quantities  $J^{\frac{1}{2}} e^{iw}$  and  $e^{-iw} J^{\frac{1}{2}}$ . Now it is easily verified from the matrix representatives that these two quantities are respectively equal to

$$\left. \begin{aligned} J^{\frac{1}{2}} e^{iw} &= e^{iw} (J + \hbar)^{\frac{1}{2}} \\ e^{-iw} J^{\frac{1}{2}} &= (J + \hbar)^{\frac{1}{2}} e^{-iw} \end{aligned} \right\} \quad (61)$$

and

and that their products in either order are

$$\begin{aligned} J^{\frac{1}{2}} e^{iw} \cdot e^{-iw} J^{\frac{1}{2}} &= J \\ e^{-iw} J^{\frac{1}{2}} \cdot J^{\frac{1}{2}} e^{iw} &= (J + \hbar)^{\frac{1}{2}} e^{-iw} \cdot e^{iw} (J + \hbar)^{\frac{1}{2}} = J + \hbar. \end{aligned}$$

These results hold in spite of the inequality in (58). They show that



when we are dealing with ordinary dynamical variables which are algebraic functions of  $q$  and  $p$  and which therefore involve the action and angle variables only through the two quantities  $J^{\frac{1}{2}}e^{i\omega}$  and  $e^{-i\omega}J^{\frac{1}{2}}$ , we may count  $e^{i\omega}$  and  $e^{-i\omega}$  as truly reciprocal quantities without getting into error. Thus we can freely use the action and angle variables in complete analogy with the classical theory without getting incorrect results.

The wave equation for the harmonic oscillator with Hamiltonian (47) is

$$i\hbar \frac{d}{dt}(q'|) = \frac{1}{2m} \left\{ -\hbar^2 \frac{\partial^2}{\partial q'^2} + m^2 \omega^2 q'^2 \right\} (q'|).$$

The wave functions representing stationary states are those periodic solutions of this equation, for which the operator  $i\hbar d/dt$  is the same as multiplication by an energy eigenvalue  $H'$  and therefore satisfy

$$H'(q'|) = \frac{1}{2m} \left\{ -\hbar^2 \frac{\partial^2}{\partial q'^2} + m^2 \omega^2 q'^2 \right\} (q'|). \quad (62)$$

The general solution of this equation has been given by Schrödinger.† It provides us with the transformation function  $(q'|H')$  connecting the  $q$ - and  $H$ -representations, one of which, it may be noted, has a discrete set of basic states while the other has a continuous range.

We shall here obtain some of the solutions representing states of lowest energy. Equation (62) reduces to

$$\left\{ \frac{d^2}{dq'^2} - \frac{q'^2}{a^4} + \frac{2n+1}{a^2} \right\} (q'|) = 0, \quad (63)$$

where  $a^2$  is the number  $\hbar/m\omega$  and  $H'$  has been put equal to  $(n+\frac{1}{2})\hbar\omega$ ,  $n$  being a positive integer or zero. Put

$$(q'|) = f(q')e^{-q'^2/2a^2}.$$

Equation (63) now becomes

$$\frac{d^2 f}{dq'^2} - 2 \frac{df}{dq'} \frac{q'}{a^2} + f \left[ \frac{q'^2}{a^4} - \frac{1}{a^2} \right] + \left[ -\frac{q'^2}{a^4} + \frac{2n+1}{a^2} \right] f = 0$$

or

$$\frac{d^2 f}{dq'^2} - 2 \frac{q'}{a^2} \frac{df}{dq'} + \frac{2n}{a^2} f = 0.$$

The solution of this equation, with  $n$  any non-negative integer, is a finite power series in  $q$ . For

$$n = 0, \quad 1, \quad 2, \quad 3, \quad \dots$$

the solutions are easily verified to be

$$f(q') = 1, \quad q', \quad q'^2 - \frac{1}{2}a^2, \quad q'^3 - \frac{3}{2}q'a^2, \quad \dots$$

† Schrödinger, *Ann. d. Physik*, 79 (1926), 514.



The successive eigenfunctions are thus

$$\left. \begin{aligned} (q'|0) &= e^{-q'^2/2a^2} & (q'|1) &= q'e^{-q'^2/2a^2} \\ (q'|2) &= (q'^2 - \tfrac{1}{2}a^2)e^{-q'^2/2a^2} & (q'|3) &= (q'^3 - \tfrac{3}{2}q'a^2)e^{-q'^2/2a^2}, \dots \end{aligned} \right\} \quad (64)$$

### 37. The Gibbs Ensemble

In our work up to the present we have been assuming all along that our dynamical system at each instant of time is in a definite state, that is to say, its motion is specified as completely and accurately as is possible without conflicting with the general principles of the theory. In the classical theory this would mean, of course, that all the coordinates and momenta have specified values. Now we may be interested in a motion which is specified to a lesser extent than this maximum possible. The present section will be devoted to the methods to be used in such a case.

The procedure in classical mechanics is to introduce what is called a *Gibbs ensemble*, the idea of which is as follows. We consider all the dynamical coordinates and momenta as Cartesian coordinates in a certain space, the *phase space*, whose number of dimensions is twice the number of degrees of freedom of the system. Any state of the system can then be represented by a point in this space. This point will move according to the classical equations of motion (11). Suppose, now, that we are not given that the system is in a definite state at any time, but only that it is in one or other of a number of possible states according to a definite probability law. We should then be able to represent it by a fluid in the phase space, the mass of fluid in any volume of the phase space being the total probability of the system being in any state whose representative point lies in that volume. Each particle of the fluid will be moving according to the equations of motion (11). If we introduce the density  $\rho$  of the fluid at any point, equal to the probability per unit volume of phase space of the system being in the neighbourhood of the corresponding state, we shall have the equation of conservation

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= - \sum_r \left\{ \frac{\partial}{\partial q_r} \left( \rho \frac{dq_r}{dt} \right) + \frac{\partial}{\partial p_r} \left( \rho \frac{dp_r}{dt} \right) \right\} \\ &= - \sum_r \left\{ \frac{\partial}{\partial q_r} \left( \rho \frac{\partial H}{\partial p_r} \right) - \frac{\partial}{\partial p_r} \left( \rho \frac{\partial H}{\partial q_r} \right) \right\} \\ &= -[\rho, H]. \end{aligned} \quad (65)$$



This may be considered as the equation of motion for the fluid, since it determines the density  $\rho$  for all time if  $\rho$  is given initially as a function of the  $q$ 's and  $p$ 's. It is, apart from the minus sign, of the same form as the ordinary equation of motion (12) for a dynamical variable.

The requirement that the total probability of the system being in any state shall be unity gives us a normalizing condition for  $\rho$

$$\iint \rho \, dq dp = 1, \quad (66)$$

the integration being over the whole of phase space and the single differential  $dq$  or  $dp$  being written to denote the product of all the  $dq$ 's or  $dp$ 's. If  $\beta$  denotes any function of the dynamical variables, the average value of  $\beta$  will be

$$\iint \beta \rho \, dq dp. \quad (67)$$

It makes only a trivial alteration in the theory, but often facilitates discussion, if we work with a density  $\rho$  differing from the above one by a positive constant factor,  $k$  say, so that we have instead of (66)

$$\iint \rho \, dq dp = k. \quad (68)$$

With this density we can picture the fluid as representing a number  $k$  of similar dynamical systems, all following through their motions independently in the same place, without any mutual disturbance or interaction. The density at any point would then be the probable or average number of systems in the neighbourhood of any state per unit volume of phase space, and expression (67) would give the average total value of  $\beta$  for all the systems. Such a set of dynamical systems, which is the ensemble introduced by Gibbs, is usually not realizable in practice, except as a rough approximation, but it forms all the same a useful theoretical abstraction.

We shall now see that there exists a corresponding density  $\rho$  in quantum mechanics, having properties analogous to the above. It was first introduced by *von Neumann*. Its existence is rather surprising in view of the fact that phase space has no meaning in quantum mechanics, there being no possibility of assigning numerical values simultaneously to the  $q$ 's and  $p$ 's.

We consider a dynamical system which is at a certain time in one or other of a number of possible states according to some given probability law. These states may be either a discrete set or a con-



tinuous range, or both together. We shall here take for definiteness the case of a discrete set and suppose them labelled by a parameter  $m$ . Let their normalized representatives in some representation be  $(\xi'|m)$  and let the probability of the system being in the  $m$ -th state be  $P_m$ . We then define the quantum density  $\rho$  through its representative:

$$(\xi'|\rho|\xi'') = \sum_m (\xi'|m)P_m(m|\xi''). \quad (69)$$

Let  $\rho'$  be any eigenvalue of  $\rho$  and  $(\xi'|)$  an eigen- $\psi$  belonging to this eigenvalue, so that

$$\int \sum_m (\xi'|m)P_m(m|\xi'') d\xi'' (\xi''|) = \rho'(\xi'|),$$

if we assume the  $\xi''$ 's to take on continuous ranges of values, for definiteness. Multiplying this equation by  $(|\xi')$ , the conjugate complex of  $(\xi'|)$ , and integrating over all  $\xi'$ , we get

$$\int \int \sum_m (|\xi') d\xi' (\xi'|m)P_m(m|\xi'') d\xi'' (\xi''|) = \rho' \int (|\xi') d\xi' (\xi'|),$$

which may be written

$$\sum_m \left| \int (|\xi') d\xi' (\xi'|m) \right|^2 P_m = \rho' \int |(\xi'|)|^2 d\xi'.$$

Now  $P_m$ , being a probability, can never be negative. It follows that  $\rho'$  cannot be negative. Thus  $\rho$  has no negative eigenvalues, in analogy with the fact that the classical density  $\rho$  is never negative.

Let us now obtain the equation of motion for our quantum  $\rho$ . The  $(\xi'|m)$ 's and  $(m|\xi'')$ 's in (69) will vary with the time in accordance with Schrödinger's wave equation (3) and its conjugate complex (4), while the  $P_m$ 's will remain constant, since the system, so long as it is left undisturbed, cannot change over from being represented by one wave function to being represented by another, so that the probability of its being represented by any particular wave function must remain constant. We thus have from (69)

$$\begin{aligned} i\hbar \frac{d}{dt} (\xi'|\rho|\xi'') &= \sum_m \left\{ i\hbar \frac{d(\xi'|m)}{dt} P_m(m|\xi'') - (\xi'|m) P_m \left[ -i\hbar \frac{d(m|\xi'')}{dt} \right] \right\} \\ &= \sum_m \int \{ (\xi'|H|\xi''') d\xi''' (\xi'''|m) P_m(m|\xi'') - \\ &\quad - (\xi'|m) P_m(m|\xi''') d\xi''' (\xi'''|H|\xi'') \} \\ &= \int \{ (\xi'|H|\xi''') d\xi''' (\xi'''|\rho|\xi'') - (\xi'|\rho|\xi''') d\xi''' (\xi'''|H|\xi'') \}, \end{aligned}$$

by using (69) again. This result may be written symbolically

$$i\hbar \dot{\rho} = H\rho - \rho H \quad (70)$$



or

$$\dot{\rho} = -[\rho, H],$$

and is thus the proper quantum analogue of the classical equation of motion (65). Our quantum  $\rho$ , like the classical one, is determined for all time if it is given initially.

From the assumption of § 12, the average value of any observable  $\beta$  when the system is in the state represented by  $(\xi'|m)$  is

$$\iint (m|\xi') d\xi' (\xi'|\beta|\xi'') d\xi'' (\xi''|m).$$

Hence if the system is distributed over the various states represented by the  $(\xi'|m)$ 's according to the probability law  $P_m$ , the average value of  $\beta$  will be

$$\begin{aligned} \sum_m \iint (m|\xi') d\xi' (\xi'|\beta|\xi'') d\xi'' (\xi''|m) P_m &= \iint (\xi'|\beta|\xi'') d\xi'' (\xi''|\rho|\xi') d\xi' \\ &= \int (\xi'|\beta\rho|\xi') d\xi' = \int (\xi'|\rho\beta|\xi') d\xi'. \end{aligned} \quad (71)$$

This is the analogue of the expression (67) of the classical theory. Whereas in the classical theory we have to multiply  $\beta$  by  $\rho$  and take the integral of the product over all phase space, in the quantum theory we have to multiply  $\beta$  by  $\rho$  and 'integrate along the diagonal' in the representative of the product. We have further, using the condition that the  $(\xi'|m)$ 's are normalized,

$$\int (\xi'|\rho|\xi') d\xi' = \sum_m \int (\xi'|m) P_m (m|\xi') d\xi' = \sum_m P_m = 1, \quad (72)$$

since the total probability of the system being in any state is unity. This is the analogue of equation (66). One more result, which follows directly from expression (35) of Chapter IV for interpreting representatives of states, is that the probability of the  $\xi$ 's having values in the neighbourhood of  $\xi'$  per unit range of the  $\xi$ 's is

$$\sum_m |(\xi'|m)|^2 P_m = (\xi'|\rho|\xi'). \quad (73)$$

This gives a physical meaning to the integrand on the left-hand side of (72).

As in the classical theory, we may take a density equal to  $k$  times the above  $\rho$  and consider it as representing a Gibbs ensemble of  $k$  similar dynamical systems, between which there is no mutual disturbance or interaction. We shall then have  $k$  on the right-hand side of (72), and (71) will give the total average  $\beta$  for all the members of the ensemble, while (73) will give the total probability of a member of the



ensemble having values for its  $\xi$ 's in the neighbourhood of  $\xi'$ , per unit range of the  $\xi''$ 's.

An important application of the foregoing theory is that it enables one to get a clearer understanding of the significance of the normalization of a  $\psi$  labelled by parameters that take on continuous ranges of values, as defined by equation (23) of Chapter IV. Let us take a system with  $n$  degrees of freedom describable in terms of canonical coordinates and momenta and suppose that it is in one or other of the simultaneous eigenstates of all the momenta, the probability of its being in an eigenstate belonging to eigenvalues for the  $p$ 's between  $p'$  and  $p' + dp'$  being  $P_{p'} dp'$ . Then in a representation in which the  $q$ 's are diagonal, the density  $\rho$  will be represented by

$$(q'|\rho|q'') = \int (q'|p') P_{p'} dp' (p'|q''). \quad (74)$$

The  $(q'|p')$ 's here are the  $q$ -representatives of the eigenstates of the momenta and are given by equation (43) of Chapter V. Thus

$$(q'|\rho|q'') = h^{-n} \int e^{i(q'_1 p'_1 + \dots + q'_n p'_n)/\hbar} P_{p'} dp' e^{-i(q''_1 p'_1 + \dots + q''_n p'_n)/\hbar}$$

and

$$(q'|\rho|q') = h^{-n} \int P_{p'} dp'. \quad (75)$$

These  $(q'|p')$ 's are normalized in accordance with the rule for  $\psi$ 's labelled by parameters that take on continuous ranges of values, i.e.

$$\int (p'|q') dq' (q'|p'') = \delta(p' - p''),$$

and not to make the corresponding  $\psi$ 's of length unity, which would require

$$\int (p'|q') dq' (q'|p') = 1.$$

In consequence our  $\rho$  does not satisfy equation (72). In fact (75) shows at once that  $\int (q'|\rho|q') dq'$  is infinite, since the integrand is independent of  $q'$ . Thus  $\rho$  should be considered as representing an ensemble of an infinite number of systems. The total probability of a member of the ensemble having its  $q$ 's in the neighbourhood of  $q'$  per unit range of the  $q''$ 's is given by (75), where it is expressed as an integral over the momentum variables. The integrand here, namely  $h^{-n} P_{p'}$ , may thus be interpreted, in a naïve way, as the probable number of systems per unit of phase space.

We can, however, get a different interpretation for  $P_{p'}$  by going back to equation (74) and replacing in it the continuous ranges of values for the  $p$ 's by discrete sets of points lying very close to one



another, in accordance with the method explained at the end of § 24. Equation (74) goes over into

$$(q' | p | q'') = \sum_{p'} (q' | p')_D P_{p'} (p' | q'')_D,$$

by the same arguments as led to (43) of § 24. Here  $P_{p'}$  appears as the probable number of systems in one of the eigenstates of momentum, these eigenstates now forming a discrete set. Comparing these two meanings for  $P_{p'}$ , we see that they will agree if we put a volume  $h^n$  of phase space equal to a discrete state. Thus *the normalization rule for the case of continuous parameters is equivalent to counting a volume  $h^n$  of phase space as having the same weight as a discrete state.*



## VII

### MOTION IN A CENTRAL FIELD OF FORCE

#### 38. Introduction of the Angular Momentum

AN atom consists of a massive positively charged nucleus together with a number of electrons moving round, under the influence of the attractive force of the nucleus and their own mutual repulsions. An exact treatment of this dynamical system would be a very difficult mathematical problem. One can, however, gain some insight into the main features of the system by making the rough approximation of regarding each electron as moving independently in a certain *central* field of force, namely that of the nucleus, assumed fixed, together with some kind of average of the forces due to the other electrons. Thus our present problem of the motion of a particle in a central field of force forms a corner-stone in the theory of the atom.

Let the Cartesian coordinates of the particle, referred to a system of axes with the centre of force as origin, be  $x, y, z$  and the corresponding components of momentum  $p_x, p_y, p_z$ . They satisfy the quantum conditions

$$[x, y] = 0 \quad [x, p_x] = 1 \quad [x, p_y] = 0,$$

etc. The Hamiltonian, with neglect of relativistic mechanics, will be of the form

$$H = 1/2m \cdot (p_x^2 + p_y^2 + p_z^2) + V, \quad (1)$$

where  $V$ , the potential energy, is a function only of  $(x^2 + y^2 + z^2)$ .

We now introduce the components of angular momentum defined, as in the classical theory, by

$$m_x = yp_z - zp_y \quad m_y = zp_x - xp_z \quad m_z = xp_y - yp_x, \quad (2)$$

or by the vector equation  $\mathbf{m} = \mathbf{x} \times \mathbf{p}$ .

From these equations we obtain at once the identity

$$m_x x + m_y y + m_z z = 0. \quad (3)$$

We must now evaluate the P.B.'s of the angular momentum components with the dynamical variables  $x, p_x$ , etc., and with each other. This we can do most conveniently with the help of the laws (4) and (5) of § 25, thus

$$\left. \begin{aligned} [m_z, x] &= [xp_y - yp_x, x] = -y[p_x, x] = y \\ [m_z, y] &= [xp_y - yp_x, y] = x[p_y, y] = -x \end{aligned} \right\} \quad (4)$$

$$[m_z, z] = [xp_y - yp_x, z] = 0, \quad (5)$$



and similarly,

$$[m_z, p_x] = p_y \quad [m_z, p_y] = -p_x \quad (6)$$

$$[m_z, p_z] = 0, \quad (7)$$

with corresponding relations for  $m_x$  and  $m_y$ . Again

$$\left. \begin{aligned} [m_y, m_z] &= [zp_x - xp_z, m_z] = z[p_x, m_z] - [x, m_z]p_z \\ &= -zp_y + yp_z = m_x \\ [m_z, m_x] &= m_y \quad [m_x, m_y] = m_z. \end{aligned} \right\} (8)$$

These results are all the same as in the classical theory. The sign in the results (4), (6), and (8) may easily be remembered from the rule that the + sign occurs when the three dynamical variables, consisting of the two in the P.B. on the left-hand side and the one forming the result on the right, are in the cyclic order ( $xyz$ ) and the - sign occurs otherwise.

From (4) and (5) we obtain

$$\begin{aligned} [m_z, x^2 + y^2 + z^2] &= x[m_z, x] + [m_z, x]x + y[m_z, y] + [m_z, y]y \\ &= xy + yx - yx - xy = 0. \end{aligned} \quad (9)$$

Similarly from (6) and (7) we find

$$[m_z, p_x^2 + p_y^2 + p_z^2] = 0. \quad (10)$$

Thus  $m_z$  commutes with  $(x^2 + y^2 + z^2)$  and with  $(p_x^2 + p_y^2 + p_z^2)$ . It therefore commutes with the Hamiltonian  $H$  which, according to (1), is a function of these two dynamical variables only. Similarly  $m_x$  and  $m_y$  commute with  $H$ . Thus *the angular momentum is a constant of the motion*, as in the classical theory.

Equations (8) may be put in the vector form

$$\mathbf{m} \times \mathbf{m} = i\hbar \mathbf{m}. \quad (11)$$

If we have several particles with angular momenta  $\mathbf{m}_1, \mathbf{m}_2, \dots$ , each of them will satisfy (11), thus

$$\mathbf{m}_r \times \mathbf{m}_r = i\hbar \mathbf{m}_r.$$

Further, any one of these angular momenta will commute with any other, so that

$$\mathbf{m}_r \times \mathbf{m}_s + \mathbf{m}_s \times \mathbf{m}_r = 0 \quad (r \neq s).$$

Hence if  $\mathbf{M} = \sum_r \mathbf{m}_r$  is the total angular momentum,

$$\begin{aligned} \mathbf{M} \times \mathbf{M} &= \sum_{rs} \mathbf{m}_r \times \mathbf{m}_s = \sum_r \mathbf{m}_r \times \mathbf{m}_r + \sum_{r < s} (\mathbf{m}_r \times \mathbf{m}_s + \mathbf{m}_s \times \mathbf{m}_r) \\ &= i\hbar \sum_r \mathbf{m}_r = i\hbar \mathbf{M}. \end{aligned}$$



This result is of the same form as (11), so that the components of the total angular momentum  $\mathbf{M}$  of any number of particles satisfy the same commutability relations as those of the angular momentum of a single particle. Thus (11) or (8) may be regarded as *the general commutability relations satisfied by any angular momentum*. They certainly hold when the angular momentum is that of a number of particles, and may be assumed to hold also for the angular momentum of a spinning body.

### 39. Properties of Angular Momentum

We shall here consider some general properties of any angular momentum  $\mathbf{m}$ . We introduce a dynamical variable  $\theta$  defined by

$$\theta = m_x^2 + m_y^2 + m_z^2,$$

having the meaning of the square of the magnitude of the vector  $\mathbf{m}$ . With the help of (8) we obtain

$$\begin{aligned} [m_x, \theta] &= [m_x, m_y^2 + m_z^2] \\ &= [m_x, m_y]m_y + m_y[m_x, m_y] + [m_x, m_z]m_z + m_z[m_x, m_z] \\ &= m_z m_y + m_y m_z - m_y m_z - m_z m_y \\ &= 0. \end{aligned}$$

Thus  $m_x$  commutes with  $\theta$ . Similarly  $m_y$  and  $m_z$  commute with  $\theta$ . We shall now assume  $\theta$  is an observable and introduce a representation in which  $\theta$  and  $m_z$  are diagonal. Since any function  $f$  of the  $m$ 's commutes with  $\theta$ , i.e.

$$\theta f - f \theta = 0,$$

its representative will satisfy

$$\theta'(\theta' m_z' | f | \theta'' m_z'') - (\theta' m_z' | f | \theta'' m_z'') \theta'' = 0,$$

$$\text{or} \quad \{\theta' - \theta''\}(\theta' m_z' | f | \theta'' m_z'') = 0,$$

so that all its matrix elements  $(\theta' m_z' | f | \theta'' m_z'')$  will vanish except those for which  $\theta' = \theta''$ , or those which are diagonal in  $\theta$ , as we may say. Thus if we express any equation between functions of the  $m$ 's in terms of representatives, the surviving matrix elements will all be diagonal in  $\theta$  and will refer to the same eigenvalue  $\theta'$  all through the equation. Hence in such equations we may count the dynamical variable  $\theta$  simply as the number  $\theta'$ .

Taking now the equation

$$\begin{aligned} (m_x + im_y)m_z - m_z(m_x + im_y) &= -i\hbar m_y - \hbar m_x \\ &= -\hbar(m_x + im_y), \end{aligned}$$

$$\text{or} \quad (m_x + im_y)m_z = (m_z - \hbar)(m_x + im_y),$$



and expressing it in terms of representatives, we get

$$(\theta' m'_z | m_x + i m_y | \theta' m''_z) m''_z = \{m'_z - \hbar\} (\theta' m'_z | m_x + i m_y | \theta' m''_z).$$

Thus all the matrix elements  $(\theta' m'_z | m_x + i m_y | \theta' m''_z)$  of the representative of  $m_x + i m_y$  vanish except those for which  $m''_z = m'_z - \hbar$ . If we now express the equation

$$\begin{aligned} (m_x + i m_y)(m_x - i m_y) &= m_x^2 + m_y^2 - i(m_x m_y - m_y m_x) \\ &= m_x^2 + m_y^2 + \hbar m_z \\ &= \theta - m_z^2 + \hbar m_z \end{aligned}$$

in terms of representatives and equate the diagonal elements on each side referring to the eigenvalues  $\theta'$ ,  $m'_z$ , so as to get

$$\sum_{m''_z} (\theta' m'_z | m_x + i m_y | \theta' m''_z) (\theta' m''_z | m_x - i m_y | \theta' m'_z) = \theta' - m'^2_z + \hbar m'_z,$$

we shall have all the terms in the sum on the left-hand side vanishing except (at most) the one for which  $m''_z = m'_z - \hbar$ . If  $m'_z - \hbar$  is not an eigenvalue of  $m_z$ , then all the terms in the sum will vanish without exception. In any case  $\theta' - m'^2_z + \hbar m'_z$  is positive or zero and, if  $m'_z - \hbar$  is not an eigenvalue of  $m_z$ , it must be zero. Thus, considering  $\theta' - m'^2_z + \hbar m'_z$  as  $\theta' + \frac{1}{4}\hbar^2 - (m'_z - \frac{1}{2}\hbar)^2$ , we can draw the conclusions that

- (i)  $\theta' + \frac{1}{4}\hbar^2$  is positive or zero,
- (ii) for any  $\theta'$  there is a minimum  $m'_z$ , satisfying
 
$$(m'_z - \frac{1}{2}\hbar)^2 = \theta' + \frac{1}{4}\hbar^2,$$
- (iii) any other  $m'_z$  for this  $\theta'$  is greater than the minimum one by an integral multiple of  $\hbar$ .

The above conclusions provide us with an example of a mathematical phenomenon which we have not met with previously, namely, that with two commuting observables, the permissible eigenvalues of one depend on what eigenvalue we assign to the other. This phenomenon may be understood as the two observables being not altogether independent, but partially functions of one another.

By a similar piece of work to the above, based on the equation

$$(m_x - i m_y)(m_x + i m_y) = \theta - m_z^2 - \hbar m_z,$$

we can draw two further conclusions, namely,

- (ii)' for any  $\theta'$  there is a maximum  $m'_z$ , satisfying
 
$$(m'_z + \frac{1}{2}\hbar)^2 = \theta' + \frac{1}{4}\hbar^2,$$
- (iii)' any other  $m'_z$  for this  $\theta'$  is less than the maximum one by an integral multiple of  $\hbar$ .



From (ii) and (ii)' it follows that the minimum value of  $m'_z$  is  $-(\theta' + \frac{1}{4}\hbar^2)^{\frac{1}{2}} + \frac{1}{2}\hbar$  and the maximum value is  $(\theta' + \frac{1}{4}\hbar^2)^{\frac{1}{2}} - \frac{1}{2}\hbar$ . The maximum value minus the minimum value, i.e.  $2(\theta' + \frac{1}{4}\hbar^2)^{\frac{1}{2}} - \hbar$ , must be an integral multiple of  $\hbar$  not less than zero. Let us introduce the new dynamical variable  $k$ , defined by

$$k + \frac{1}{2}\hbar = (m_x^2 + m_y^2 + m_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}}, \quad (12)$$

the positive square root being taken on the right-hand side, in accordance with the definition at the end of § 11. This equation, it may be noted, gives

$$k(k + \hbar) = m_x^2 + m_y^2 + m_z^2.$$

We now have

$$k = (\theta + \frac{1}{4}\hbar^2)^{\frac{1}{2}} - \frac{1}{2}\hbar,$$

so that the eigenvalues of  $k$  are integral or half odd integral multiples of  $\hbar$  not less than zero. For each eigenvalue  $k'$  of  $k$  the eigenvalues of  $m_z$  are

$$k', k' - \hbar, k' - 2\hbar, \dots, -k' + \hbar, -k'.$$

From symmetry,  $m_x$  and  $m_y$  have the same eigenvalues as  $m_z$ .

The dynamical variable  $k$  defined by (12) is the convenient one to use for describing the magnitude of an angular momentum vector  $\mathbf{m}$ . It is preferable to the square root of  $\theta$  on account of its simpler eigenvalues. The eigenvalues of  $k$  and of the components  $m_x, m_y, m_z$  are always integral or half odd integral numbers of quanta  $\hbar$ . *If, however,  $\mathbf{m}$  is the angular momentum of a particle moving in an orbit, then the eigenvalues of  $m_x, m_y, m_z$  and  $k$  must all be integral numbers of quanta.* To verify this we take  $m_z$ , which is now of the form given by the third of equations (2), and put it in a representation in which the coordinates  $x, y, z$  are diagonal. According to § 27 its representative will be the differential operator

$$-i\hbar\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)$$

when operating to the right. The easiest way of obtaining the eigenvalues of this operator is to transform it to the cylindrical coordinates  $z, \rho, \varphi$ , in which  $\rho$  and  $\varphi$  are defined by  $x = \rho \cos \varphi$ ,  $y = \rho \sin \varphi$ . It then becomes simply  $-i\hbar \partial/\partial \varphi$ . The eigenfunctions of this operator are obviously of the form  $ae^{in\varphi}$ , where  $a$  is a function of  $z$  and  $\rho$  only and  $n$  is an integer.† The corresponding eigenvalue for  $m_z$  is  $n\hbar$ , and is thus an integral multiple of  $\hbar$ . The eigenvalues of  $k$  must then also be integral multiples of  $\hbar$ .

† It is a general requirement of our theory of representations that representatives of  $\psi$ 's and  $\phi$ 's shall always be single-valued. Thus our eigenfunctions must be single-valued.



Although the angular momentum of orbital motion of a particle must have integral eigenvalues, there is no reason why the spin angular momentum of a particle should not have half odd integral eigenvalues, since a spin angular momentum is not expressible in terms of coordinates and momenta in the form (2). In fact experimental evidence shows that electrons and many kinds of atomic nuclei have spin angular momenta with half odd integral eigenvalues. A further remark about spin is that a spin angular momentum may have a magnitude  $k$  with only one eigenvalue. This is possible since  $k$  commutes with the three components of the angular momentum, so that we do not get any inconsistency by putting  $k$  equal to a number. (This would not be possible for the  $k$  of an orbital angular momentum, since for such a  $k$  there would exist the variables  $x, p_x$ , etc., with which it does not commute.) It is found that all the more elementary particles of atomic physics, such as electrons and atomic nuclei, do have spin angular momenta with magnitudes  $k$  with only one eigenvalue. For the spin of electrons,  $k$  has the one eigenvalue  $\frac{1}{2}\hbar$ . This results in the components  $m_x, m_y, m_z$  of the spin angular momentum of an electron each having the eigenvalues  $\frac{1}{2}\hbar, -\frac{1}{2}\hbar$ . These components are thus of the form of  $\frac{1}{2}\hbar$  times the  $\sigma$ 's of § 19, since these  $\sigma$ 's each have the eigenvalues 1,  $-1$  and their commutability relations, namely equations (53) of § 19, are the same as (8), apart from the factors  $\frac{1}{2}\hbar$ . Theoretical reasons for this particular spin angular momentum for an electron will be obtained in Chapter XII. For the spins of the other elementary particles (except photons) there is at present no theoretical information, and one has to depend entirely on experimental evidence.

The components  $m_x, m_y, m_z$  of an angular momentum in different directions do not commute with each other, so that one cannot in general assign numerical values to them simultaneously. One can at most give a numerical value to the component in one particular direction. The state of the system is then said to be *spacially quantized* in that direction. There is, however, one special case in which one can assign numerical values to all the components simultaneously, namely, one can give them all the value zero, since this will not contradict the commutability relations (8). The resulting state of zero angular momentum, with  $k = 0$ , is one that is spacially quantized simultaneously in all directions and is, according to the work at the end of § 29, spherically symmetrical.



### 40. Transition to Polar Coordinates

For further discussion of the problem of motion in a central field of force it is convenient to introduce polar dynamical variables. We introduce first the radius  $r$ , defined as the positive square root

$$r = (x^2 + y^2 + z^2)^{\frac{1}{2}}.$$

If we evaluate its P.B.'s with  $p_x, p_y$ , and  $p_z$ , we obtain, with the help of formula (38) of Chapter V,

$$[r, p_x] = \frac{\partial r}{\partial x} = \frac{x}{r} \quad [r, p_y] = \frac{y}{r} \quad [r, p_z] = \frac{z}{r},$$

the same as in the classical theory. We introduce also the dynamical variable  $p_r$  defined by

$$p_r = r^{-1}(xp_x + yp_y + zp_z - i\hbar). \quad (13)$$

Its P.B. with  $r$  is given by

$$\begin{aligned} r[r, p_r] &= [r, rp_r] = [r, xp_x + yp_y + zp_z] \\ &= x[r, p_x] + y[r, p_y] + z[r, p_z] \\ &= x \cdot x/r + y \cdot y/r + z \cdot z/r = r. \end{aligned}$$

Hence

$$[r, p_r] = 1$$

or

$$rp_r - p_r r = i\hbar.$$

We can now see that  $p_r$  is real, since its conjugate complex  $\bar{p}_r$  is given by

$$\begin{aligned} \bar{p}_r &= (p_x x + p_y y + p_z z + i\hbar)r^{-1} \\ &= (xp_x + yp_y + zp_z - 2i\hbar)r^{-1} \\ &= (rp_r - i\hbar)r^{-1} = p_r. \end{aligned}$$

The commutability relation between  $r$  and  $p_r$  is just the one for a pair of canonically conjugate variables, namely equation (12) of Chapter V. Now the eigenvalues of  $r$ , from its definition as a positive square root, must be all positive or zero, so that we have obtained a contradiction to the result, proved at the end of § 26, that a dynamical variable can have a canonical conjugate only if its eigenvalues include all numbers from  $-\infty$  to  $\infty$ . This inconsistency arises from the fact that the dynamical variable  $p_r$  defined by (13) does not strictly exist, since  $r$  has the eigenvalue zero so that  $r^{-1}$  does not strictly exist. In spite of this defect the dynamical variable  $p_r$  is a useful one for the study of motion in a central field of force. Our equations, which will often involve  $p_r$  and will sometimes involve  $r^{-1}$  in other ways than through  $p_r$ , will be liable to be inaccurate, but only in so far as they



apply to the one point  $r = 0$ . It will be necessary to make a special investigation of solutions of the wave equation obtained with the help of polar variables to see whether they are satisfactory at the point  $r = 0$ . We shall do this later in this section. It was mentioned at the end of § 26 that  $e^{cq}$  is inadmissible as an operator in quantum mechanics if  $c$  is real and the eigenvalues of  $q$  extend from  $-\infty$  to  $\infty$ , but since the eigenvalues of  $r$  extend only from 0 to  $\infty$ ,  $e^{cr}$  is admissible if  $c$  is negative.

We can easily verify that our two new dynamical variables  $r$  and  $p_r$  commute with the angular momentum. Equation (9) shows us that  $m_z$  commutes with  $r^2$ . It must therefore commute also with  $r$ , since  $r$  is defined as a square-root function so that, from a theorem on page 56, everything that commutes with  $r^2$  commutes also with  $r$ . Again, for  $p_r$  we have

$$\begin{aligned} r[p_r, m_z] &= [rp_r, m_z] = [xp_x + yp_y, m_z] \\ &= -yp_x - xp_y + xp_y + yp_x = 0. \end{aligned}$$

Thus  $r$  and  $p_r$  commute with  $m_z$ , and hence also with  $m_x$  and  $m_y$  and with  $k$ .

We can now express the Hamiltonian in terms of our radial dynamical variables  $r$  and  $p_r$  and also  $k$ . We have, if  $\sum_{xyz}$  denotes a sum over cyclic permutations of the suffixes  $x, y, z$ ,

$$\begin{aligned} k(k+\hbar) &= \sum_{xyz} m_z^2 = \sum_{xyz} (xp_y - yp_x)^2 \\ &= \sum_{xyz} (xp_y xp_y + yp_x yp_x - xp_y yp_x - yp_x xp_y) \\ &= \sum_{xyz} (x^2 p_y^2 + y^2 p_x^2 - xp_x p_y y - yp_y p_x x + x^2 p_x^2 - xp_x p_x x - \\ &\hspace{15em} - 2i\hbar xp_x) \\ &= (x^2 + y^2 + z^2)(p_x^2 + p_y^2 + p_z^2) - \\ &\hspace{10em} - (xp_x + yp_y + zp_z)(p_x x + p_y y + p_z z + 2i\hbar) \\ &= r^2(p_x^2 + p_y^2 + p_z^2) - (rp_r + i\hbar)rp_r \\ &= r^2(p_x^2 + p_y^2 + p_z^2) - r^2 p_r^2. \end{aligned}$$

Hence 
$$H = \frac{1}{2m} \left( p_r^2 + \frac{k(k+\hbar)}{r^2} \right) + V. \tag{14}$$

This form for  $H$  is such that  $k$  commutes not only with  $H$ , as is necessary since  $k$  is a constant of the motion, but also with every dynamical variable occurring in  $H$ , namely both  $r$  and  $p_r$ . Thus in dealing with the Hamiltonian in this form we can treat  $k$  as a number. The permissible numbers we can take for  $k$  are its eigenvalues and are



thus positive integral multiples of  $\hbar$  or zero. The equation for the representatives of the stationary states will now read†

$$\left\{ \frac{1}{2m} \left( -\hbar^2 \frac{\partial^2}{\partial r^2} + \frac{k(k+\hbar)}{r^2} \right) + V \right\} (r|) = H'(r|), \quad (15)$$

the single variable  $r$  in the wave function  $(r|)$  being sufficient when  $k$  is counted as a number. Any value of the parameter  $H'$  for which this equation, with a permissible value for  $k$ , has a solution (satisfying the boundary conditions to be discussed later) is a possible energy-level of the system. The energy-levels (except those for which  $k = 0$ ) each belong to several independent stationary states, corresponding to the various possible eigenvalues of a Cartesian component of the angular momentum. The number of these states, for any value of  $k$ , is the odd number  $(2k/\hbar + 1)$ .

If we write down the equation for the representatives of the stationary states in the original Cartesian coordinates  $x, y, z$ , we shall have

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + V \right\} (xyz|) = H'(xyz|), \quad (16)$$

where  $\nabla^2$  is the Laplacian operator  $\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ . This becomes, on transforming to polar coordinates  $r, \theta, \phi$ ,

$$\left\{ -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right) + V \right\} (r\theta\phi|) = H'(r\theta\phi|).$$

The solutions of this equation are of the form

$$(r\theta\phi|) = \chi(r) S_n(\theta\phi),$$

where  $S_n$  is a spherical harmonic of order  $n$ , satisfying

$$\left( \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right) S_n(\theta\phi) = -n(n+1) S_n(\theta\phi),$$

$n$  being an integer, and  $\chi(r)$  is a function of  $r$  only, satisfying

$$\left\{ -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{n(n+1)}{r^2} \right) + V \right\} \chi(r) = H' \chi(r). \quad (17)$$

This equation, like (15), is such that the values of  $H'$  for which it has a solution are the energy-levels of the system.

The equivalence of equations (15) and (17) may be seen from the

† We are here omitting the primes from the variables in the wave functions. This is often convenient when one can do so without confusion, it being understood that the variables in the wave function, or in any representative, denote eigenvalues of observables and not the observables themselves.



fact that if in (15) we put  $(r|) = r\chi(r)$  we obtain just equation (17) with  $n = k/\hbar$ . The fact that the two eigenfunctions  $(r|)$  and  $\chi(r)$  are not identical but differ by this factor  $r$  is due to their different physical interpretations. A solution  $(r|)$  of (15) represents a state for which the probability of the particle lying in the spherical shell between  $r$  and  $r + dr$  is proportional to  $|(r)|^2 dr$ . On the other hand, a solution  $(xyz|)$  of (16) represents a state for which the probability of the particle lying in a small volume  $dxdydz$  is  $|(xyz)|^2 dxdydz$  or  $|\chi(r)S_n(\theta\phi)|^2 dxdydz$ , so that the probability of its lying in the spherical shell between  $r$  and  $r + dr$  is proportional to  $|\chi(r)|^2 r^2 dr$ . Thus the physical interpretations require  $(r|)$  to be proportional to  $r\chi(r)$ .

It should be noticed that not every solution of (17), when multiplied by the appropriate spherical harmonic, will give a solution of (16), as it may fail to satisfy (16) at the origin. One can see most clearly how this comes about by considering the special case for which the potential  $V$  vanishes, giving us the problem of the free particle. If we further take  $H' = 0$ , equation (16) reduces to

$$\nabla^2(xyz|) = 0 \quad (18)$$

and equation (17) to

$$\left\{ \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{n(n+1)}{r^2} \right\} \chi(r) = 0. \quad (19)$$

Now a solution of (19) for  $n = 0$  is  $\chi(r) = 1/r$ , but this solution multiplied by the appropriate spherical harmonic  $S_0 = 1$  does not satisfy (18), since, although  $\nabla^2(1/r)$  vanishes for any finite value of  $r$ , its integral through any volume about the origin is  $4\pi$ , and hence

$$\nabla^2(1/r) = 4\pi\delta(x)\delta(y)\delta(z).$$

Thus the solution  $\chi(r) = 1/r$  of (19) does not represent a stationary state of the system. Again the solution  $\chi(r) = 1/r^2$  of (19) for  $n = 1$ , when multiplied by the spherical harmonic  $S_1 = \cos\theta$ , gives a wave function  $(xyz|)$ , the integral of the square of whose modulus over any volume, however small, that contains the origin is infinite. This wave function must represent a state for which the particle is certainly at the origin and this cannot be a stationary state of zero energy for the problem of the free particle. Similarly for arbitrary  $n$  in equation (19), of the two solutions  $\chi(r) = r^n$  and  $\chi(r) = r^{-n-1}$ , the second will not give the representative of a stationary state of the system.

It thus appears that equation (17) is not adequate to replace equation (16) as the necessary and sufficient condition for the representa-



tive of a stationary state. Equation (17) must be supplemented by a suitable boundary condition at the point  $r = 0$ . Any solution  $\chi(r)$  of (17) for which the integral  $\int_0 r^2 |\chi(r)|^2 dr$  is not convergent must certainly be rejected, and also some for which this integral is convergent, namely those which, when operated on by  $\nabla^2$ , give an infinite result involving the  $\delta$  function at the origin. These conditions show that only those solutions are to be allowed which, if they tend to infinity as  $r \rightarrow 0$ , do so more slowly than  $1/r$ . The corresponding boundary condition for the function  $(r|)$  of equation (15) is that it shall tend to zero as  $r \rightarrow 0$ .

There are also boundary conditions for the eigenfunction at  $r = \infty$ . If we are interested only in 'closed' states, i.e. states for which the particle does not go off to infinity, we must restrict the integral  $\int_0^\infty |(r)|^2 dr$  or  $\int_0^\infty r^2 |\chi(r)|^2 dr$  to be convergent. These closed states, however, are not the only ones that are physically permissible, as we can also have states in which the particle arrives from infinity, is scattered by the central field of force, and goes off to infinity again. For these states the wave function  $(xyz|)$  may remain finite as  $r \rightarrow \infty$ . Such states will be dealt with in Chapter IX under the heading of collision problems. In any case the eigenfunction  $(xyz|)$  must not tend to infinity as  $r \rightarrow \infty$ , or it will represent a state that has no physical meaning.

#### 41. Energy-levels of the Hydrogen Atom

The above analysis may be applied to the problem of the hydrogen atom with neglect of the relativistic variation of mass with velocity and the spin of the electron. The potential energy  $V$  is now†  $-e^2/r$ , so that equation (15) becomes

$$\left\{ \frac{d^2}{dr^2} - \frac{k(k+1)}{r^2} + \frac{2me^2}{\hbar^2} \frac{1}{r} \right\} (r|) = -\frac{2mH'}{\hbar^2} (r|), \quad (20)$$

when written in terms of a new dynamical variable  $k$ , equal to  $\hbar^{-1}$  times the previous  $k$ . A thorough investigation of this equation has been given by Schrödinger.‡ We shall here obtain its eigenvalues  $H'$  from a consideration of its eigenfunctions expressed in the form of power series.

† The  $e$  here, denoting minus the charge on an electron, is, of course, to be distinguished from the  $e$  denoting the base of exponentials.

‡ Schrödinger, *Ann. d. Physik*, **79** (1926), 361.



It is convenient to put

$$(r|) = f(r)e^{-r/a}, \quad (21)$$

introducing the new function  $f(r)$ , where  $a$  is one or other of the square roots

$$a = \pm\sqrt{-\hbar^2/2mH'}. \quad (22)$$

Equation (20) now becomes

$$\left\{ \frac{d^2}{dr^2} - \frac{2}{a} \frac{d}{dr} - \frac{k(k+1)}{r^2} + \frac{2me^2}{\hbar^2} \frac{1}{r} \right\} f(r) = 0. \quad (23)$$

We look for a solution of this equation in the form of a power series

$$f(r) = \sum_s c_s r^s, \quad (24)$$

in which consecutive values for  $s$  differ by unity although these values themselves need not be integers. On substituting (24) in (23) we obtain

$$\sum_s c_s \{s(s-1)r^{s-2} - (2s/a)r^{s-1} - k(k+1)r^{s-2} + (2me^2/\hbar^2)r^{s-1}\} = 0,$$

which gives, on equating to zero the coefficient of  $r^{s-2}$ , the following relation between successive coefficients  $c_s$ ,

$$c_s[s(s-1) - k(k+1)] = c_{s-1}[2(s-1)/a - 2me^2/\hbar^2]. \quad (25)$$

We saw in the preceding section that only those eigenfunctions  $(r|)$  are allowed that tend to zero with  $r$  and hence, from (21),  $f(r)$  must tend to zero with  $r$ . The series (24) must therefore terminate on the side of small  $s$  and the minimum value of  $s$  must be greater than zero. Now the only possible minimum values of  $s$  are those that make the coefficient of  $c_s$  in (25) vanish, i.e.  $k+1$  and  $-k$ , and the second of these is negative or zero. Thus the minimum value of  $s$  must be  $k+1$ . Since  $k$  is always an integer, the values of  $s$  will all be integers. The series (24) will in general extend to infinity on the side of large  $s$ . For large values of  $s$  the ratio of successive terms is

$$\frac{c_s}{c_{s-1}} r = \frac{2r}{sa}$$

according to (25). Thus the series (24) will always converge, as the ratios of the higher terms to one another are the same as for the series

$$\sum_s \frac{1}{s!} \left( \frac{2r}{a} \right)^s, \quad (26)$$

which converges to  $e^{2r/a}$ .

We must now examine how our solution  $(r|)$  behaves for large values of  $r$ . We must distinguish between the two cases of  $H'$  positive



and  $H'$  negative. For  $H'$  negative,  $a$  given by (22) will be real. Suppose we take the positive value for  $a$ . Then as  $r \rightarrow \infty$  the sum of the series (24) will tend to infinity according to the same law as the sum of the series (26), i.e. the law  $e^{2r/a}$ . Thus, from (21),  $(r|)$  will tend to infinity according to the law  $e^{r/a}$  and will not represent a physically possible state. There is therefore in general no permissible solution of (20) for negative values of  $H'$ . An exception arises, however, whenever the series (24) terminates on the side of large  $s$ , in which case the boundary conditions are all satisfied. The condition for this termination of the series is that the coefficient of  $c_{s-1}$  in (25) shall vanish for some value of the suffix  $s-1$  not less than its minimum value  $k+1$ , which is the same as the condition that

$$\frac{s}{a} - \frac{me^2}{\hbar^2} = 0$$

for some integer  $s$  not less than  $k+1$ . With the help of (22) this condition becomes

$$H' = -\frac{me^4}{2s^2\hbar^2}, \quad (27)$$

and is thus a condition for the energy-level  $H'$ . Since  $s$  may be any positive integer, the formula (27) gives a discrete set of negative energy-levels for the hydrogen atom. These are in agreement with experiment. Each of them (except the lowest one  $s = 1$ ) may occur with various possible values for  $k$ , namely, any positive or zero integer less than  $s$ . This multiplicity is in addition to that mentioned in the preceding section arising from the various possible values for a component of angular momentum, which multiplicity occurs with any central field of force. The  $k$  multiplicity occurs only with an inverse square law of force and even then is removed when one takes relativistic mechanics into account, as will be found in Chapter XII. The solution of (20) when  $H'$  satisfies (27) tends to zero exponentially as  $r \rightarrow \infty$  and thus represents a closed state (corresponding to an elliptic orbit in Bohr's theory).

For any positive values of  $H'$ ,  $a$  given by (22) will be pure imaginary. The series (24), which is roughly the same as the series (26), will now have a sum that remains finite as  $r \rightarrow \infty$ . Thus  $(r|)$  given by (21) will now remain finite as  $r \rightarrow \infty$  and will therefore be a permissible solution of (20), since it will correspond to an eigenfunction  $(xyz|)$  that tends to zero according to the law  $1/r$  as  $r \rightarrow \infty$ . Hence in addition to the discrete set of negative energy-levels (27), all positive energy-



levels are allowed. The states of positive energy are not closed, since their representatives  $(r|)$  do not make the integral  $\int_0^\infty |(r|)|^2 dr$  converge. (These states correspond to the hyperbolic orbits of Bohr's theory.)

## 42. Selection Rules

If a dynamical system is set up in a certain stationary state, it will remain in that stationary state so long as it is not acted upon by outside forces. Any atomic system in practice, however, frequently gets acted upon by external electromagnetic fields, under whose influence it is liable to cease to be in one stationary state and to make a transition to another. The theory of such transitions will be developed in §§ 47 and 48. A result of this theory is that, to a high degree of accuracy, transitions between two states cannot occur under the influence of electromagnetic radiation if, in a Heisenberg representation with these two stationary states as two of the basic states, the matrix element, referring to these two states, of the representative of the total electric displacement  $\mathbf{D}$  of the system vanishes. Now it happens for many atomic systems that the great majority of the matrix elements of  $\mathbf{D}$  in a Heisenberg representation do vanish, and hence there are severe limitations on the possibilities for transitions. The rules that express these limitations are called *selection rules*.

The idea of selection rules can be refined by a more detailed application of the results of the theory of § 48, according to which the matrix elements of the different Cartesian components of the vector  $\mathbf{D}$  are associated with different states of polarization of the electromagnetic radiation. The nature of this association is just what one would get if one considered the matrix elements, or rather their real parts, as the amplitudes of harmonic oscillators which interact with the field of radiation according to classical electrodynamics. We shall consider some examples to illustrate this.

There is a general method for obtaining all selection rules, as follows. Let us call the constants of the motion which are diagonal in our Heisenberg representation  $\alpha$ 's and let  $D$  be one of the Cartesian components of  $\mathbf{D}$ . We must obtain an algebraic equation connecting  $D$  and the  $\alpha$ 's which does not involve any dynamical variables other than  $D$  and the  $\alpha$ 's and which is linear in  $D$ . Such an equation will be of the form

$$\sum_r f_r D g_r = 0, \quad (28)$$

where the  $f_r$ 's and  $g_r$ 's are functions of the  $\alpha$ 's only. If this equa-



tion is expressed in terms of representatives, it gives us

$$\sum_r f_r(\alpha') (\alpha' | D | \alpha'') g_r(\alpha'') = 0,$$

or 
$$(\alpha' | D | \alpha'') \sum_r f_r(\alpha') g_r(\alpha'') = 0,$$

which shows that  $(\alpha' | D | \alpha'') = 0$  unless

$$\sum_r f_r(\alpha') g_r(\alpha'') = 0. \quad (29)$$

This last equation, giving the connexion which must exist between  $\alpha'$  and  $\alpha''$  in order that  $(\alpha' | D | \alpha'')$  may not vanish, constitutes the selection rule, so far as the component  $D$  of  $\mathbf{D}$  is concerned.

Our work on the harmonic oscillator in § 36, in connexion with equations (52) and (53) there, provides an example of a selection rule. If the harmonic oscillator carries an electric charge, its electric displacement  $D$  will be proportional to  $q$ . The selection rule is then given by equation (53) there, and is that only those transitions can take place in which the energy  $H$  changes by a single quantum  $\hbar\omega$ .

We shall now obtain the selection rules for  $m_z$  and  $k$  for an electron moving in a central field of force. The components of electric displacement are here proportional to the Cartesian coordinates  $x, y, z$ . Taking first  $m_z$ , we have that  $m_z$  commutes with  $z$ , or that

$$m_z z - z m_z = 0.$$

This is an equation of the required type (28), giving us the selection rule

$$m'_z - m''_z = 0$$

for the  $z$ -component of the displacement. Again, from equations (4) we have

$$[m_z, [m_z, x]] = [m_z, y] = -x$$

or 
$$m_z^2 x - 2m_z x m_z + x m_z^2 - \hbar^2 x = 0,$$

which is also of the type (28) and gives us the selection rule

$$m_z'^2 - 2m'_z m_z'' + m_z''^2 - \hbar^2 = 0$$

or 
$$(m'_z - m_z'' - \hbar)(m'_z - m_z'' + \hbar) = 0$$

for the  $x$ -component of the displacement. The selection rule for the  $y$ -component is the same. Thus our selection rules for  $m_z$  are that *in transitions associated with radiation with a polarization corresponding to an electric dipole in the  $z$ -direction,  $m'_z$  cannot change, while in transitions associated with a polarization corresponding to an electric dipole in the  $x$ -direction or  $y$ -direction,  $m'_z$  must change by  $\pm\hbar$ .*

We can determine more accurately the state of polarization of the



radiation associated with a transition in which  $m'_z$  changes by  $\pm\hbar$ , by considering the condition for the non-vanishing of matrix elements of  $x+iy$  and  $x-iy$ . We have

$$[m_z, x+iy] = y-ix = -i(x+iy)$$

$$\text{or} \quad m_z(x+iy) - (x+iy)(m_z+\hbar) = 0,$$

which is again of the type (28). It gives

$$m'_z - m''_z - \hbar = 0$$

as the condition that  $(m'_z|x+iy|m''_z)$  shall not vanish. Similarly,

$$m'_z - m''_z + \hbar = 0$$

is the condition that  $(m'_z|x-iy|m''_z)$  shall not vanish. Hence

$$(m'_z|x-iy|m'_z-\hbar) = 0$$

$$\text{or} \quad (m'_z|x|m'_z-\hbar) = i(m'_z|y|m'_z-\hbar) = (a+ib)e^{i\omega t}$$

say,  $a$ ,  $b$ , and  $\omega$  being real, and similarly

$$(m'_z-\hbar|x|m'_z) = -i(m'_z-\hbar|y|m'_z) = (a-ib)e^{-i\omega t}.$$

Thus the vector  $\frac{1}{2}\{(m'_z|\mathbf{D}|m'_z-\hbar) + (m'_z-\hbar|\mathbf{D}|m'_z)\}$ , which determines the state of polarization of the radiation associated with transitions for which  $m''_z = m'_z - \hbar$ , has the following three components

$$\left. \begin{aligned} & \frac{1}{2}\{(m'_z|x|m'_z-\hbar) + (m'_z-\hbar|x|m'_z)\} \\ & \quad = \frac{1}{2}\{(a+ib)e^{i\omega t} + (a-ib)e^{-i\omega t}\} = a \cos \omega t - b \sin \omega t \\ & \frac{1}{2}\{(m'_z|y|m'_z-\hbar) + (m'_z-\hbar|y|m'_z)\} \\ & \quad = \frac{1}{2}i\{-(a+ib)e^{i\omega t} + (a-ib)e^{-i\omega t}\} = a \sin \omega t + b \cos \omega t \\ & \frac{1}{2}\{(m'_z|z|m'_z-\hbar) + (m'_z-\hbar|z|m'_z)\} = 0. \end{aligned} \right\} \quad (30)$$

From the form of these components we see that the associated radiation moving in the  $z$ -direction will be circularly polarized, that moving in any direction in the  $xy$  plane will be linearly polarized in this plane, and that moving in intermediate directions will be elliptically polarized. The direction of circular polarization for radiation moving in the  $z$ -direction will depend on whether  $\omega$  is positive or negative, and this will depend on which of the two states  $m'_z$  or  $m''_z = m'_z - \hbar$  has the greater energy.

We shall now determine the selection rule for  $k$ . We have

$$\begin{aligned} [k(k+\hbar), z] &= [m_x^2, z] + [m_y^2, z] \\ &= -ym_x - m_x y + xm_y + m_y x \\ &= 2(m_y x - m_x y + i\hbar z) \\ &= 2(m_y x - ym_x) = 2(xm_y - m_x y). \end{aligned}$$



Similarly,  $[k(k+\hbar), x] = 2(y m_z - m_y z)$

and  $[k(k+\hbar), y] = 2(m_x z - x m_z).$

Hence

$$\begin{aligned}
 [k(k+\hbar), [k(k+\hbar), z]] &= 2[k(k+\hbar), m_y x - m_x y + i\hbar z] \\
 &= 2m_y[k(k+\hbar), x] - 2m_x[k(k+\hbar), y] + 2i\hbar[k(k+\hbar), z] \\
 &= 4m_y(y m_z - m_y z) - 4m_x(m_x z - x m_z) + 2\{k(k+\hbar)z - zk(k+\hbar)\} \\
 &= 4(m_x x + m_y y + m_z z)m_z - 4(m_x^2 + m_y^2 + m_z^2)z + \\
 &\quad + 2\{k(k+\hbar)z - zk(k+\hbar)\}.
 \end{aligned}$$

The first term here vanishes, from (3), leaving us with

$$[k(k+\hbar), [k(k+\hbar), z]] = -2\{k(k+\hbar)z + zk(k+\hbar)\},$$

which gives

$$\begin{aligned}
 k^2(k+\hbar)^2z - 2k(k+\hbar)zk(k+\hbar) + zk^2(k+\hbar)^2 - \\
 - 2\hbar^2\{k(k+\hbar)z + zk(k+\hbar)\} = 0. \quad (31)
 \end{aligned}$$

Similar equations hold for  $x$  and  $y$ . These equations are of the required type (28), and give us the selection rule

$$\begin{aligned}
 k'^2(k'+\hbar)^2 - 2k'(k'+\hbar)k''(k''+\hbar) + k''^2(k''+\hbar)^2 - \\
 - 2\hbar^2k'(k'+\hbar) - 2\hbar^2k''(k''+\hbar) = 0,
 \end{aligned}$$

which reduces to

$$(k' + k'' + 2\hbar)(k' + k'')(k' - k'' + \hbar)(k' - k'' - \hbar) = 0.$$

A transition can take place between two states  $k'$  and  $k''$  only if one of these four factors vanishes.

Now the first of the factors,  $(k' + k'' + 2\hbar)$ , can never vanish, since the eigenvalues of  $k$  are all positive or zero. The second,  $(k' + k'')$ , can vanish only if  $k' = 0$  and  $k'' = 0$ . But transitions between two states with these values for  $k$  cannot occur on account of the selection rule for  $m_z$ , as may be seen from the following argument. If two states (labelled respectively with a single prime and a double prime) are such that  $k' = 0$  and  $k'' = 0$ , then, according to the discussion at the end of § 39, each Cartesian component of the angular momentum must vanish for each of them, i.e.  $m'_x = m'_y = m'_z = 0$  and  $m''_x = m''_y = m''_z = 0$ . The selection rule for  $m_z$  now shows that the matrix elements of  $x$  and  $y$  referring to the two states must vanish, as the value of  $m_z$  does not change during the transition, and the similar selection rule for  $m_x$  or  $m_y$  shows that the matrix element of  $z$  also vanishes. Thus



transitions between the two states cannot occur. Our selection rule for  $k$  now reduces to

$$(k' - k'' + \hbar)(k' - k'' - \hbar) = 0,$$

showing that  $k$  must change by  $\pm \hbar$ . This selection rule may be written

$$k'^2 - 2k'k'' + k''^2 - \hbar^2 = 0,$$

and since this is the condition that a matrix element  $(k'|z|k'')$  shall not vanish, we get the equation

$$k^2 z - 2kzk + zk^2 - \hbar^2 z = 0$$

$$\text{or} \quad [k, [k, z]] = -z, \quad (32)$$

a result which could not easily be obtained in a more direct way.

### 43. The Zeeman Effect for the Hydrogen Atom

We shall now consider the system of a hydrogen atom in a uniform magnetic field. The Hamiltonian (1) with  $V = -e^2/r$ , which describes the hydrogen atom in no external field, gets modified by the magnetic field, the modification, according to classical mechanics, consisting in the replacement of the components of momentum,  $p_x, p_y, p_z$ , by  $p_x + e/c \cdot A_x, p_y + e/c \cdot A_y, p_z + e/c \cdot A_z$ , where  $A_x, A_y, A_z$  are the components of the vector potential describing the field. For a uniform field of magnitude  $\mathcal{H}$  in the direction of the  $z$ -axis we may take  $A_x = -\frac{1}{2}\mathcal{H}y, A_y = \frac{1}{2}\mathcal{H}x, A_z = 0$ . The classical Hamiltonian will then be

$$H = \frac{1}{2m} \left\{ \left( p_x - \frac{1}{2} \frac{e}{c} \mathcal{H} y \right)^2 + \left( p_y + \frac{1}{2} \frac{e}{c} \mathcal{H} x \right)^2 + p_z^2 \right\} - \frac{e^2}{r}.$$

This classical Hamiltonian may be taken over into the quantum theory if we add on to it a term giving the effect of the spin of the electron. According to experimental evidence and according to the theory of Chapter XII, the electron has a magnetic moment  $-e\hbar/2mc \cdot \sigma$ , where  $\sigma$  is a vector with the properties given in § 19. The energy of this magnetic moment in the magnetic field will be  $e\hbar\mathcal{H}/2mc \cdot \sigma_z$ . Thus the total quantum Hamiltonian will be

$$H = \frac{1}{2m} \left\{ \left( p_x - \frac{1}{2} \frac{e}{c} \mathcal{H} y \right)^2 + \left( p_y + \frac{1}{2} \frac{e}{c} \mathcal{H} x \right)^2 + p_z^2 \right\} - \frac{e^2}{r} + \frac{e\hbar\mathcal{H}}{2mc} \sigma_z. \quad (33)$$

There ought strictly to be other terms in this Hamiltonian giving the interaction of the magnetic moment of the electron with the electric field of the nucleus of the atom, but this effect is small, of the same order of magnitude as that of the relativistic variation of the mass



of the electron with its velocity, and will be neglected here. It will be taken into account in the relativistic theory of the electron given in Chapter XII.

If the magnetic field is not too large, we can neglect terms involving  $\mathcal{H}^2$ , so that the Hamiltonian (33) reduces to

$$\begin{aligned} H &= \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) - \frac{e^2}{r} + \frac{e\mathcal{H}}{2mc}(xp_y - yp_x) + \frac{e\hbar\mathcal{H}}{2mc}\sigma_z \\ &= \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) - \frac{e^2}{r} + \frac{e\mathcal{H}}{2mc}(m_z + \hbar\sigma_z). \end{aligned} \quad (34)$$

The extra terms due to the magnetic field are now  $e\mathcal{H}/2mc \cdot (m_z + \hbar\sigma_z)$ . But these extra terms commute with the total Hamiltonian and are thus constants of the motion. This makes the problem very easy. The stationary states of the system, i.e. the eigenstates of the Hamiltonian (34), will be those eigenstates of the Hamiltonian for no field that are simultaneously eigenstates of the observables  $m_z$  and  $\sigma_z$ , or at least of the one observable  $m_z + \hbar\sigma_z$ , and the energy-levels of the system will be those for the system with no field, given by (27) if one considers only closed states, increased by an eigenvalue of  $e\mathcal{H}/2mc \cdot (m_z + \hbar\sigma_z)$ . Thus stationary states of the system with no field which are spacially quantized in the  $z$ -direction (i.e. for which  $m_z$  has the numerical value  $m'_z$ , an integral multiple of  $\hbar$ ) and for which also  $\sigma_z$  has the numerical value  $\sigma'_z = \pm 1$ , will still be stationary states when the field is applied. Their energy will be increased by an amount consisting of the sum of two parts, a part  $e\mathcal{H}/2mc \cdot m'_z$  arising from the orbital motion, which part may be considered as due to an orbital magnetic moment  $-em'_z/2mc$ , and a part  $e\mathcal{H}/2mc \cdot \hbar\sigma'_z$  arising from the spin. The ratio of the orbital magnetic moment to the orbital angular momentum  $m'_z$  is  $-e/2mc$ , which is half the ratio of the spin magnetic moment to the spin angular momentum. This fact is sometimes referred to as the magnetic anomaly of the spin.

Since the energy-levels now involve  $m_z$ , the selection rule for  $m_z$  obtained in the preceding section becomes capable of direct comparison with experiment. We take a Heisenberg representation in which, among other constants of the motion,  $m_z$  and  $\sigma_z$  are diagonal. The selection rule for  $m_z$  now requires  $m_z$  to change by  $\hbar$ , 0, or  $-\hbar$ , while  $\sigma_z$ , since it commutes with the electric displacement, will not change at all. Thus the energy difference between the two states taking part in the transition process will differ by an amount



$e\hbar\mathcal{H}/2mc$ , 0, or  $-e\hbar\mathcal{H}/2mc$  from its value for no magnetic field. Hence, from Bohr's frequency condition, the frequency of the associated electromagnetic radiation will differ by  $e\mathcal{H}/4\pi mc$ , 0, or  $-e\mathcal{H}/4\pi mc$  from that for no magnetic field. This means that each spectral line for no magnetic field gets split up by the field into three components. If one considers radiation moving in the  $z$ -direction, then from (30) the two outer components will be circularly polarized, while the central undisplaced one will be of zero intensity. These results are in agreement with experiment and also with the classical theory of the Zeeman effect. The agreement with the classical theory ceases, however, when one takes into account relativistic mechanics and the interaction of the spin with the electric field of the nucleus.

#### 44. Combination of Angular Momenta

Suppose we have two particles moving in the central field of force, having as angular momenta the vectors  $\mathbf{m}$  and  $\boldsymbol{\mu}$ . We can introduce the dynamical variables  $k$  and  $\kappa$ , defined by (12) and

$$\kappa + \frac{1}{2}\hbar = (\mu_x^2 + \mu_y^2 + \mu_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}},$$

respectively, to describe the magnitudes of these vectors. The total angular momentum will then be the vector  $\mathbf{M} = \mathbf{m} + \boldsymbol{\mu}$ , whose magnitude  $K$  is defined by

$$K + \frac{1}{2}\hbar = (M_x^2 + M_y^2 + M_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}}.$$

Each of the dynamical variables  $k$  and  $\kappa$  commutes with all the components of  $\mathbf{m}$ ,  $\boldsymbol{\mu}$ , and  $\mathbf{M}$ . Thus  $k$ ,  $\kappa$ ,  $K$  will commute with each other and can be given numerical values simultaneously. Our problem now is to determine the possible numerical values for  $K$  when  $k$  and  $\kappa$  have given numerical values.

The easiest way of solving this problem is to suppose  $k$  and  $\kappa$  are equal to two given numbers, as we can do since they commute with all the dynamical variables mentioned in the problem, and then to use a matrix representation in which  $m_z$  and  $\mu_z$  are diagonal. We can ignore all dynamical variables describing the dynamical system that are not functions of the components of  $\mathbf{m}$  and  $\boldsymbol{\mu}$ . Our matrix representation will then have only a finite number of rows and columns, each labelled by a number  $m'_z$  having one of the values  $k, k-\hbar, k-2\hbar, \dots, -k$  and a number  $\mu'_z$  having one of the values  $\kappa, \kappa-\hbar, \kappa-2\hbar, \dots, -\kappa$ . The possible values of  $M'_z = m'_z + \mu'_z$  will then be  $k+\kappa, k+\kappa-\hbar, k+\kappa-2\hbar, \dots, -k-\kappa$ . The number of times each of



them occurs is given by the following scheme (if we assume for definiteness that  $k \geq \kappa$ ),

$$\left. \begin{array}{ccccccc} k+\kappa, & k+\kappa-\hbar, & k+\kappa-2\hbar, & \dots, & k-\kappa, & k-\kappa-\hbar, & \dots \\ 1 & 2 & 3 & \dots & 2\kappa+1 & 2\kappa+1 & \dots \\ & & & \dots & -k+\kappa, & -k+\kappa-\hbar, & \dots, & -k-\kappa \\ & & & \dots & 2\kappa+1 & 2\kappa & \dots & 1 \end{array} \right\} \quad (35)$$

If we now make a transformation to a representation in which  $K$  and  $M_z$  are diagonal, the number of rows and columns of the matrices for which  $M_z$  has a given value  $M'_z$  must remain unaltered. If  $K', K'', \dots$  are the possible values for  $K$ , there will be a set of rows and columns having the  $M_z$ -values  $K', K'-\hbar, \dots, -K'$ , another set having the  $M_z$ -values  $K'', K''-\hbar, \dots, -K''$ , etc. Comparing this distribution of  $M_z$ -values with (35), we see that the possible values for  $K$  must be

$$k+\kappa, k+\kappa-\hbar, k+\kappa-2\hbar, \dots, k-\kappa. \quad (36)$$

This result is a quite general one applying to the combination of any two angular momenta, not necessarily the orbital angular momenta of two particles. For example, it could be applied to the orbital angular momentum and spin of an electron. In this case, since the spin angular momentum has the magnitude  $\kappa = \frac{1}{2}\hbar$ , it shows that when the orbital angular momentum has the magnitude  $k$ , the combined angular momentum can have only one or other of the two magnitudes  $k \pm \frac{1}{2}\hbar$ .

We now have a general method for dealing with complicated atomic systems. For an isolated system the total angular momentum  $\mathbf{M}$  is always a constant of the motion, and its magnitude  $K$  together with one of its components  $M_z$  will be two commuting constants of the motion. We try to express  $\mathbf{M}$  as the sum of two angular momenta  $\mathbf{m}$  and  $\boldsymbol{\mu}$  whose magnitudes  $k$  and  $\kappa$  are constants of the motion. If we can do this, then we try to express either of the parts,  $\mathbf{m}$  say, itself as a sum of two angular momenta,  $\mathbf{m}_1$  and  $\mathbf{m}_2$  say, whose magnitudes  $k_1$  and  $k_2$  are constants of the motion, and so on. We obtain in this way a series of constants of the motion  $M_z, K, k, \kappa, k_1, k_2, \dots$ , which all commute with each other and may, if there are enough of them, be taken as defining a Heisenberg representation. The possible numerical values for the  $K, k, \kappa, \dots$  specifying a row and column are restricted by the general rule (36). The energy will be some function of  $K, k, \kappa, k_1, k_2, \dots$ , but independent of  $M_z$ . In general one cannot



secure that  $k, \kappa, k_1, k_2$  are exactly constants of the motion, but one may be able to choose them so that they are approximately so and then apply a perturbation method, as discussed in the next chapter.

We shall now obtain the selection rule for the magnitude  $K$  of the total angular momentum  $\mathbf{M}$  of a general atomic system. Let  $\mathbf{m}$  be the orbital angular momentum of one of the electrons, whose coordinates are  $x, y, z$  say, and let  $\mathbf{M} - \mathbf{m} = \boldsymbol{\mu}$ . It is not necessary for the present discussion that the magnitudes  $k$  and  $\kappa$  of the two angular momenta  $\mathbf{m}$  and  $\boldsymbol{\mu}$  into which  $M$  has been split up should be constants of the motion. We must obtain the condition that the  $(K', K'')$  matrix element of  $x, y$ , or  $z$  shall not vanish. This is evidently the same as the condition that the  $(K', K'')$  matrix element of  $\lambda_1, \lambda_2$ , or  $\lambda_3$  shall not vanish, where  $\lambda_1, \lambda_2$  and  $\lambda_3$  are any three independent linear functions of  $x, y$  and  $z$  with numerical coefficients, or more generally with any coefficients that commute with  $K$  and are thus represented by matrices which are diagonal with respect to  $K$ . Let

$$\lambda_0 = M_x x + M_y y + M_z z$$

$$\lambda_x = M_y z - M_z y - i\hbar x$$

$$\lambda_y = M_z x - M_x z - i\hbar y$$

$$\lambda_z = M_x y - M_y x - i\hbar z.$$

We have

$$\begin{aligned} M_x \lambda_x + M_y \lambda_y + M_z \lambda_z &= \sum_{xyz} (M_x M_y z - M_x M_z y - i\hbar M_x x) \\ &= \sum_{xyz} (M_x M_y - M_y M_x - i\hbar M_z) z = 0 \end{aligned} \quad (37)$$

from the general condition (11) for angular momentum. Thus  $\lambda_x, \lambda_y$  and  $\lambda_z$  are not linearly independent functions of  $x, y$  and  $z$ . Any two of them, however, together with  $\lambda_0$  are three linearly independent functions of  $x, y$  and  $z$  and may be taken as the above  $\lambda_x, \lambda_y$  and  $\lambda_z$ , since the coefficients  $M_x, M_y, M_z$  all commute with  $K$ . Our problem thus reduces to finding the condition that the  $(K', K'')$  matrix elements of  $\lambda_0, \lambda_x, \lambda_y$  and  $\lambda_z$  shall not vanish. The physical meanings of these  $\lambda$ 's are that  $\lambda_0$  is proportional to the component of the vector  $(x, y, z)$  in the direction of the vector  $\mathbf{M}$ , and  $\lambda_x, \lambda_y, \lambda_z$  are proportional to the Cartesian components of the component of  $(x, y, z)$  perpendicular to  $\mathbf{M}$ .

From (4) together with the condition that  $x, y$  and  $z$  commute with  $\boldsymbol{\mu}$  we obtain

$$\left. \begin{aligned} [M_z, x] &= [m_z + \mu_z, x] = y \\ [M_z, y] &= -x \quad [M_z, z] = 0. \end{aligned} \right\} \quad (38)$$



Hence

$$\begin{aligned} [M_z, \lambda_0] &= [M_z, M_x]x + M_x[M_z, x] + [M_z, M_y]y + M_y[M_z, y] \\ &= M_yx + M_xy - M_xy - M_yx = 0. \end{aligned}$$

Thus  $\lambda_0$  commutes with  $M_z$ , and from symmetry it must commute also with  $M_x$  and  $M_y$ , so that it must commute with  $K$ . It follows that only the diagonal elements  $(K'|\lambda_0|K')$  of  $\lambda_0$  can differ from zero, so the selection rule is that  $K$  cannot change so far as this component of the electric displacement is concerned.

With further applications of (38) we obtain

$$\begin{aligned} [M_z, \lambda_x] &= [M_z, M_y]z - M_z[M_z, y] - i\hbar[M_z, x] \\ &= -M_xz + M_zx - i\hbar y = \lambda_y \\ [M_z, \lambda_y] &= M_z[M_z, x] - [M_z, M_x]z - i\hbar[M_z, y] \\ &= M_zy - M_yz + i\hbar x = -\lambda_x \\ [M_z, \lambda_z] &= [M_z, M_x]y + M_x[M_z, y] - [M_z, M_y]x - M_y[M_z, x] \\ &= M_yy - M_xx + M_xx - M_yy = 0. \end{aligned}$$

These relations between  $M_z$  and  $\lambda_x, \lambda_y, \lambda_z$  are of exactly the same form as the relations (4), (5) between  $m_z$  and  $x, y, z$ , and also (37) is of the same form as (3). The dynamical variables  $\lambda_x, \lambda_y, \lambda_z$  thus have the same properties relative to the angular momentum  $\mathbf{M}$  as  $x, y, z$  have relative to  $\mathbf{m}$ . The deduction in § 42 of the selection rule for  $k$  when the electric displacement is proportional to  $(x, y, z)$  can therefore be taken over and applied to the selection rule for  $K$  when the electric displacement is proportional to  $(\lambda_x, \lambda_y, \lambda_z)$ . We find in this way that, so far as  $\lambda_x, \lambda_y, \lambda_z$  are concerned, the selection rule for  $K$  is that it must change by  $\pm\hbar$ .

Collecting results, we have as the selection rule for  $K$  that it must change by 0 or  $\pm\hbar$ . We have considered the electric displacement produced by only one of the electrons, but the same selection rule must hold for each electron and thus also for the total electric displacement.



## VIII

### PERTURBATION THEORY

#### 45. General Remarks

IN the preceding two chapters exact treatments were given of some simple dynamical systems in the quantum theory. Most quantum problems, however, cannot be solved exactly with the present resources of mathematics, as they lead to equations whose solutions cannot be expressed in finite terms with the help of the ordinary functions of analysis. For such problems one can often use a perturbation method. This consists in splitting up the Hamiltonian into two parts, one of which must be simple and the other small. The first part may then be considered as the Hamiltonian of a simplified or unperturbed system, which can be dealt with exactly, and the addition of the second will then require small corrections, of the nature of a perturbation, in the solution for the unperturbed system. The requirement that the first part shall be simple requires in practice that it shall not involve the time explicitly. If the second part contains a small numerical factor  $\epsilon$ , we can obtain the solution of our equations for the perturbed system in the form of a power series in  $\epsilon$ , which, provided it converges, will give the answer to our problem with any desired accuracy. Even when the series does not converge, the first approximation obtained by means of it is usually fairly accurate.

There are two distinct methods in perturbation theory. In one of these the perturbation is considered as causing *a modification of the states* of the unperturbed system (with the space-time meaning of 'states'). In the other we do not consider any modification to be made in the states of the unperturbed system, but we suppose that the perturbed system, instead of remaining permanently in *one* of these states, is continually changing from one to another, or *making transitions*, under the influence of the perturbation. Which method is to be used in any particular case depends on the nature of the problem to be solved. The first method is useful usually only when the perturbing energy (the correction in the Hamiltonian for the undisturbed system) does not involve the time explicitly, and is then applied to the stationary states. It can be used for calculating things that do not refer to any definite time, such as the energy-levels of the stationary



states of the perturbed system, or, in the case of collision problems, the probability of scattering through a given angle. The second method must, on the other hand, be used for solving all problems involving a consideration of time, such as those about the transient phenomena that occur when the perturbation is suddenly applied, or more generally problems in which the perturbation varies with the time in any way (i.e. in which the perturbing energy involves the time explicitly in an arbitrary way). Again, this second method must be used in collision problems, even though the perturbing energy does not here involve the time explicitly, if one wishes to calculate absorption and emission probabilities, since these probabilities, unlike a scattering probability, cannot be defined without reference to a state of affairs that varies with the time.

#### 46. The Change in the Energy-levels caused by a Perturbation

The first of the above-mentioned methods will now be applied to the calculation of the changes in the energy-levels of a system caused by a perturbation. We assume the perturbing energy, like the Hamiltonian for the unperturbed system, not to involve the time explicitly. Our problem has a meaning, of course, only provided the energy-levels of the unperturbed system are discrete and the differences between them are large compared with the changes in them caused by the perturbation. This fact results in the treatment of perturbation problems by the first method having some different features according to whether the energy-levels of the unperturbed system are discrete or continuous.

Let the Hamiltonian of the perturbed system be

$$H = H_0 + V, \quad (1)$$

$H_0$  being the Hamiltonian of the unperturbed system and  $V$  the small perturbing energy. By hypothesis each eigenvalue  $H'$  of  $H$  lies very close to one and only one eigenvalue  $H''_0$  of  $H_0$ . We shall use the same number of primes to specify any eigenvalue of  $H$  and the eigenvalue of  $H_0$  to which it lies very close. Thus we shall have  $H''$  differing from  $H''_0$  by a small quantity of order  $V$  and differing from  $H'_0$  by a quantity that is not small unless  $H'_0 = H''_0$ . We must now take care always to use different numbers of primes to specify eigenvalues of  $H$  and  $H_0$  which we do not want to lie very close together.



We have to solve the equation

$$H\psi = \{H_0 + V\}\psi = H'\psi$$

$$\text{or} \quad \{H' - H_0\}\psi = V\psi. \quad (2)$$

Let  $\psi_0$  be an eigen- $\psi$  of  $H_0$  belonging to the eigenvalue  $H'_0$  and suppose the  $\psi$  and  $H'$  that satisfy (2) to differ from  $\psi_0$  and  $H'_0$  only by small quantities and to be expressed as

$$\left. \begin{aligned} \psi &= \psi_0 + \psi_1 + \psi_2 + \dots \\ H' &= H'_0 + a_1 + a_2 + \dots, \end{aligned} \right\} (3)$$

where  $\psi_1$  and  $a_1$  are of the first order of smallness (i.e. the same order as  $V$ ),  $\psi_2$  and  $a_2$  are of the second order, and so on. Substituting these expressions in (2), we obtain

$$\{H'_0 - H_0 + a_1 + a_2 + \dots\}\{\psi_0 + \psi_1 + \psi_2 + \dots\} = V\{\psi_0 + \psi_1 + \dots\}.$$

If we now separate the terms of zero order, of the first order, of the second order, and so on, we get the following set of equations,

$$\left. \begin{aligned} \{H'_0 - H_0\}\psi_0 &= 0 \\ \{H'_0 - H_0\}\psi_1 + a_1\psi_0 &= V\psi_0 \\ \{H'_0 - H_0\}\psi_2 + a_1\psi_1 + a_2\psi_0 &= V\psi_1 \\ &\dots \end{aligned} \right\} (4)$$

The first of these equations tells us, what we have already assumed, that  $\psi_0$  is an eigen- $\psi$  of  $H_0$  belonging to the eigenvalue  $H'_0$ . The others enable us to calculate the various corrections  $\psi_1, \psi_2, \dots, a_1, a_2, \dots$ .

For the further discussion of these equations it is convenient to introduce a representation in which  $H_0$  is diagonal, i.e. a Heisenberg representation for the unperturbed system, and to take  $H_0$  itself as one of the observables whose eigenvalues label the representatives. Let the others, in the event of others being necessary, as is the case when there is more than one eigenstate of  $H_0$  belonging to any eigenvalue, be called  $\beta$ 's. The representatives of  $\psi, \psi_0, \psi_1, \psi_2$  are then  $(H''_0\beta''|)$ ,  $(H''_0\beta''|0)$ ,  $(H''_0\beta''|1)$ ,  $(H''_0\beta''|2)$  respectively. Since  $\psi_0$  is an eigen- $\psi$  of  $H_0$  belonging to the eigenvalue  $H'_0$ , we have

$$(H''_0\beta''|0) = \delta_{H''_0 H'_0}(\beta''|0), \quad (5)$$

where  $(\beta''|0)$  is some function of the variables  $\beta''$ . With the help of this result the second of equations (4), written in terms of representatives, becomes

$$\{H'_0 - H'_0\}(H''_0\beta''|1) + a_1\delta_{H''_0 H'_0}(\beta''|0) = \sum_{\beta'} (H''_0\beta''|V|H'_0\beta')(\beta'|0). \quad (6)$$



Putting  $H''_0 = H'_0$  here, we get

$$a_1(\beta''|0) = \sum_{\beta'} (H'_0\beta''|V|H'_0\beta')(\beta'|0). \quad (7)$$

Equation (7) is of the form of the standard equation in the theory of eigenvalues, so far as the variables  $\beta'$  are concerned. It shows that the various possible values for  $a_1$  are the eigenvalues of the matrix  $(H'_0\beta''|V|H'_0\beta')$ . This matrix is a part of the representative of the perturbing energy in the Heisenberg representation for the unperturbed system, namely, the part consisting of those elements that refer to the same unperturbed energy-level  $H'_0$  for their row and column. Each of these values for  $a_1$  gives, to the first order, an energy-level of the perturbed system lying close to the energy-level  $H'_0$  of the unperturbed system.† There may thus be several energy-levels of the perturbed system lying close to the one energy-level  $H'_0$  of the unperturbed system, their number being anything not exceeding the number of independent states of the unperturbed system belonging to the energy-level  $H'_0$ . In this way the perturbation may cause a separation or partial separation of the energy-levels that coincide at  $H'_0$  for the unperturbed system.

Equation (7) also determines, to the zero order, the representatives  $(H''_0\beta''|0)$  of the stationary states of the perturbed system belonging to energy-levels lying close to  $H'_0$ , any solution  $(\beta'|0)$  of (7) substituted in (5) giving one such representative. Each of these stationary states of the perturbed system approximates to one of the stationary states of the unperturbed system, but the converse, that each stationary state of the unperturbed system approximates to one of the stationary states of the perturbed system, is not true, since the general stationary state of the unperturbed system belonging to the energy-level  $H'_0$  is represented by the right-hand side of (5) with an arbitrary function  $(\beta''|0)$ . The problem of finding which stationary states of the unperturbed system approximate to stationary states of the perturbed system, i.e. the problem of finding the solutions  $(\beta'|0)$  of (7), corresponds to the problem of secular perturbations in classical mechanics. It should be noted that the above results are independent of the values of all those matrix elements of the

† To distinguish these energy-levels one from another we should require some more elaborate notation, since according to the present notation they must all be specified by the same number of primes, namely, by the number of primes specifying the energy-level of the unperturbed system from which they arise. For our present purposes, however, this more elaborate notation is not required.



perturbing energy which refer to two different energy-levels  $H'_0$  and  $H''_0$  of the unperturbed system.

Let us see what the above results become in the specially simple case when there is only one stationary state of the unperturbed system belonging to each energy-level.<sup>†</sup> In this case  $H_0$  alone fixes the representation, no  $\beta$ 's being required. The sum in (7) now reduces to a single term and we get

$$a_1 = (H'_0|V|H'_0). \quad (8)$$

There is only one energy-level of the perturbed system lying close to any energy-level of the unperturbed system and *the change in energy is equal, in the first order, to the corresponding diagonal element of the perturbing energy in the Heisenberg representation for the unperturbed system, or to the average value of the perturbing energy for the corresponding unperturbed state.* The latter formulation of the result is the same as in classical mechanics when the unperturbed system is multiply periodic.

We shall proceed to calculate the second-order correction  $a_2$  in the energy-level for the case when the unperturbed system is non-degenerate. Equation (5) for this case reads

$$(H''_0|0) = \delta_{H''_0 H'_0},$$

with neglect of an unimportant numerical factor, and equation (6) reads

$$\{H'_0 - H''_0\}(H''_0|1) + a_1 \delta_{H''_0 H'_0} = (H''_0|V|H'_0).$$

This gives us a value for  $(H''_0|1)$  when  $H''_0 \neq H'_0$ , namely

$$(H''_0|1) = \frac{(H''_0|V|H'_0)}{H'_0 - H''_0}. \quad (9)$$

The third of equations (4), written in terms of representatives, becomes

$$\{H'_0 - H''_0\}(H''_0|2) + a_1(H''_0|1) + a_2 \delta_{H''_0 H'_0} = \sum_{H'''_0} (H''_0|V|H'''_0)(H'''_0|1).$$

Putting  $H''_0 = H'_0$  here, we get

$$a_1(H'_0|1) + a_2 = \sum_{H'''_0} (H'_0|V|H'''_0)(H'''_0|1),$$

which reduces, with the help of (8), to

$$a_2 = \sum_{H''_0 \neq H'_0} (H'_0|V|H''_0)(H''_0|1).$$

<sup>†</sup> A system with only one stationary state belonging to each energy-level is often called *non-degenerate* and one with two or more stationary states belonging to an energy-level is called *degenerate*, although these words are not very appropriate from the modern point of view.



Substituting for  $(H_0''|1)$  from (9), we obtain finally

$$a_2 = \sum_{H_0'' \neq H_0'} \frac{(H_0'|V|H_0'')(H_0''|V|H_0')}{H_0' - H_0''},$$

giving for the total energy change to the second order

$$a_1 + a_2 = (H_0'|V|H_0') + \sum_{H_0'' \neq H_0'} \frac{(H_0'|V|H_0'')(H_0''|V|H_0')}{H_0' - H_0''}. \quad (10)$$

The method may be developed for the calculation of the higher approximations if required. General recurrence formulas giving the  $n$ -th order corrections in terms of those of lower order have been obtained by Born, Heisenberg, and Jordan.†

#### 47. The Perturbation considered as causing Transitions

We shall now consider the second of the two perturbation methods mentioned in § 45. We suppose again that we have an unperturbed system governed by a Hamiltonian  $H_0$  which does not involve the time explicitly, and a perturbing energy  $V$  which can now be an arbitrary function of the time. The Hamiltonian for the perturbed system is again  $H = H_0 + V$ . For the present method it does not make any essential difference whether the energy-levels of the unperturbed system, i.e. the eigenvalues of  $H_0$ , form a discrete or continuous set. We shall, however, take the discrete case, for definiteness.

We shall again work with a Heisenberg representation for the unperturbed system, but as there will now be no advantage in taking  $H_0$  itself as one of the observables whose eigenvalues label the representatives, we shall suppose we have a general set of  $\alpha$ 's to label the representatives. The representative of  $H_0$  will be diagonal, of the form

$$(\alpha'|H_0|\alpha'') = H_0' \delta_{\alpha'\alpha''}, \quad (11)$$

like (14) of § 32. We shall have to make use of both the representations considered at the end of § 32, differing one from the other with regard to the phase factors and fixed in the Heisenberg and Schrödinger pictures respectively. Equation (11) holds for both. As in § 32, we shall use stars to distinguish representatives in the representation which is fixed in the Schrödinger picture. The two representatives of a  $\psi$  are connected by

$$(\alpha'|)^* = e^{-iH_0't/\hbar}(\alpha'|), \quad (12)$$

like equation (17) of § 32. The Schrödinger wave equation, which

† *Z. f. Physik*, 35 (1925), 565.



holds with the starred representatives, reads

$$\begin{aligned} i\hbar \frac{d}{dt}(\alpha'|)^* &= \sum_{\alpha''} (\alpha'|H_0 + V|\alpha'')^*(\alpha'')^* \\ &= H'_0(\alpha'|)^* + \sum_{\alpha''} (\alpha'|V|\alpha'')^*(\alpha'')^*. \end{aligned} \quad (13)$$

The representative  $(\alpha'|)$  of a state will not satisfy the Schrödinger equation, but will satisfy instead the following equation, obtained by substituting (12) in (13),

$$\begin{aligned} i\hbar \left[ -iH'_0 \hbar^{-1} e^{-iH'_0 t/\hbar} (\alpha'|) + e^{-iH'_0 t/\hbar} \frac{d}{dt} (\alpha'|) \right] \\ = H'_0 e^{-iH'_0 t/\hbar} (\alpha'|) + \sum_{\alpha''} (\alpha'|V|\alpha'')^* e^{-iH'_0 t/\hbar} (\alpha''). \end{aligned}$$

This reduces to

$$\begin{aligned} i\hbar \frac{d}{dt} (\alpha'|) &= \sum_{\alpha''} e^{i(H'_0 - H''_0)t/\hbar} (\alpha'|V|\alpha'')^* (\alpha'') \\ &= \sum_{\alpha''} (\alpha'|V|\alpha'') (\alpha''). \end{aligned} \quad (14)$$

The representative  $(\alpha'|V|\alpha'')^*$  of the perturbing energy  $V$  does not depend on  $t$ , except in so far as  $V$  itself involves  $t$  explicitly, but the representative  $(\alpha'|V|\alpha'')$  appearing in our equation (14) varies rapidly with  $t$ , according to the Heisenberg law  $e^{i(H'_0 - H''_0)t/\hbar}$  when one neglects the explicit dependence of  $V$  on  $t$ .

Equation (14) is the fundamental equation of the present method in perturbation theory. It is an exact equation, no use having yet been made of the fact that the perturbation is small. It shows how the representative of a state of a perturbed system varies with the time when the representation is chosen so that the whole of this variation is caused by the perturbation, and thus expresses most clearly the way in which the perturbation may be considered as causing a continual change in the state of the system. At any instant the probability of the  $\alpha$ 's having specified values  $\alpha'$  is

$$P' = |(\alpha'|)^*|^2 = |(\alpha')|^2 \quad (15)$$

provided  $P'$  is normalized.

We shall now obtain an approximate solution to equation (14) for a given initial value of the representative  $(\alpha'|)$  of the state. Since  $V$  is small, the rate of change of  $(\alpha'|)$  is small and  $(\alpha'|)$  remains approximately equal to its initial value, at any rate for times



that do not differ too much from the initial time. We can thus obtain a first approximation by substituting for  $(\alpha''|)$  in the right-hand side of (14) its initial value and then performing a simple integration. We may then obtain a second approximation by substituting the first approximation in the right-hand side of (14), and so on indefinitely.

Let the initial value of  $(\alpha'|)$ , i.e. the value at time  $t = 0$ , be  $a_0(\alpha')$ , or  $a'_0$  say, for brevity. We shall then have in the first approximation for the value of  $(\alpha'|)$  at an arbitrary time  $\tau$ ,

$$\begin{aligned} (\alpha'|)_{\tau} &= a'_0 - i\hbar^{-1} \sum_{\alpha''} \int_0^{\tau} (\alpha'|V|\alpha'')_t a''_0 dt \\ &= a'_0 + a'_{1\tau} \end{aligned}$$

say,  $a'_{1\tau}$  being the first-order correction, whose value at time  $\tau$  is

$$a'_{1\tau} = -i\hbar^{-1} \sum_{\alpha''} a''_0 \int_0^{\tau} (\alpha'|V|\alpha'')_t dt. \quad (16)$$

The second approximation at an arbitrary time  $T$  will now be

$$\begin{aligned} (\alpha'|)_{T} &= a'_0 - i\hbar^{-1} \sum_{\alpha''} \int_0^T (\alpha'|V|\alpha'')_{\tau} [a''_0 + a''_{1\tau}] d\tau \\ &= a'_0 + a'_{1T} + a'_{2T}, \end{aligned}$$

where  $a_2$ , the second-order correction, has the value at time  $T$

$$\begin{aligned} a'_{2T} &= -i\hbar^{-1} \sum_{\alpha''} \int_0^T (\alpha'|V|\alpha'')_{\tau} a''_{1\tau} d\tau \\ &= -\hbar^{-2} \sum_{\alpha''\alpha'''} a'''_0 \int_0^T (\alpha'|V|\alpha'')_{\tau} d\tau \int_0^{\tau} (\alpha''|V|\alpha''')_t dt. \end{aligned} \quad (17)$$

The probability (15) of the  $\alpha$ 's having the values  $\alpha'$  at any time is now, to the second order of accuracy,

$$\begin{aligned} P' &= (a'_0 + a'_1 + a'_2)(\bar{a}'_0 + \bar{a}'_1 + \bar{a}'_2) \\ &= a'_0 \bar{a}'_0 + (a'_1 \bar{a}'_0 + a'_0 \bar{a}'_1) + (a'_2 \bar{a}'_0 + a'_1 \bar{a}'_1 + a'_0 \bar{a}'_2) + \dots \\ &= P'_0 + P'_1 + P'_2 + \dots, \end{aligned} \quad (18)$$

$P'_0$  being the initial value of this probability and  $P'_1$  and  $P'_2$  being the first- and second-order corrections.

Suppose now that we are given, not the initial value  $a'_0$  of  $(\alpha'|)$ , but



only the initial probability  $P'_0$  of the  $\alpha$ 's having any specified values  $\alpha'$ , and want to calculate the probability at any subsequent time of the  $\alpha$ 's having specified values. We now know only the modulus of  $\alpha'_0$  and not its phase, so that we must average over all phases. This averaging results in a considerable simplification in the expression (18) for  $P'$ , since this expression is bilinear in  $a_0$  and  $\bar{a}_0$  [both  $a_1$  and  $a_2$  being linear functions of  $a_0$  according to (16) and (17)], and thus consists of a sum of terms of the form  $a''_0 \bar{a}'''_0$ . The average of  $a''_0 \bar{a}'''_0$  or  $a_0(\alpha'') \bar{a}_0(\alpha''')$  will vanish except when  $\alpha''' = \alpha''$ , so that the only surviving terms will be those of the form  $a''_0 \bar{a}''_0$ . In this way  $P'_1$  at time  $\tau$  reduces to

$$\begin{aligned} P'_{1\tau} &= a'_{1\tau} \bar{a}'_0 + a'_0 \bar{a}'_{1\tau} \\ &= \left[ -i\hbar^{-1} a'_0 \int_0^\tau (\alpha' | V | \alpha')_t dt \right] \bar{a}'_0 + a'_0 \left[ i\hbar^{-1} \bar{a}'_0 \int_0^\tau (\alpha' | V | \alpha')_t dt \right] \\ &= 0. \end{aligned}$$

Similarly,  $P'_2$  at time  $T$  reduces to

$$\begin{aligned} P'_{2T} &= a'_{2T} \bar{a}'_0 + a'_{1T} \bar{a}'_{1T} + a'_0 \bar{a}'_{2T} \\ &= -\hbar^{-2} a'_0 \bar{a}'_0 \sum_{\alpha''} \int_0^T (\alpha' | V | \alpha'')_\tau d\tau \int_0^\tau (\alpha'' | V | \alpha')_t dt + \\ &\quad + \hbar^{-2} \sum_{\alpha''} a''_0 \bar{a}''_0 \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2 - \\ &\quad - \hbar^{-2} a'_0 \bar{a}'_0 \sum_{\alpha''} \int_0^T (\alpha'' | V | \alpha')_\tau d\tau \int_0^\tau (\alpha' | V | \alpha'')_t dt, \end{aligned}$$

use being made, in dealing with the third term, of the fact that the matrix  $(\alpha' | V | \alpha'')$  is Hermitian. If we interchange  $t$  and  $\tau$  in this third term, we can combine it with the first term to give

$$\begin{aligned} &-\hbar^{-2} |a'_0|^2 \sum_{\alpha''} \left[ \int_0^T d\tau \int_0^\tau dt + \int_0^T dt \int_0^t d\tau \right] (\alpha' | V | \alpha'')_\tau (\alpha'' | V | \alpha')_t \\ &= -\hbar^{-2} |a'_0|^2 \sum_{\alpha''} \int_0^T d\tau \int_0^T dt (\alpha' | V | \alpha'')_\tau (\alpha'' | V | \alpha')_t \\ &= -\hbar^{-2} |a'_0|^2 \sum_{\alpha''} \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2. \end{aligned}$$



Thus our expression for  $P'_2$  becomes

$$\begin{aligned} P'_{2T} &= \hbar^{-2} \sum_{\alpha''} \{ |a''_0|^2 - |a'_0|^2 \} \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2 \\ &= \hbar^{-2} \sum_{\alpha''} \{ P''_0 - P'_0 \} \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2, \end{aligned}$$

and the probability  $P'$  of the  $\alpha$ 's having the values  $\alpha'$  is, to the second order of accuracy,

$$P'_T = P'_0 + \hbar^{-2} \sum_{\alpha''} \{ P''_0 - P'_0 \} \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2. \quad (19)$$

This result is capable of a simple interpretation. If we suppose that initially the  $\alpha$ 's certainly have the values  $\alpha'' \neq \alpha'$ , so that  $P''_0 = 1$ ,  $P'_0 = 0$  for  $\alpha'' \neq \alpha'$  (in which special case the averaging over the phases of the  $a_0$ 's is not necessary), then the right-hand side of (19) reduces to the single term

$$\hbar^{-2} \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2 = P(\alpha'' \alpha') \quad (20)$$

say. This may be interpreted as the probability of the system making a transition from the state  $\alpha''$  to the state  $\alpha'$  under the influence of the perturbation  $V$  during the interval of time 0 to  $T$ . It is symmetrical between  $\alpha'$  and  $\alpha''$ . Returning now to the general case, we see that (19) may be regarded as expressing that the change in the probability of the  $\alpha$ 's having the values  $\alpha'$  during the time interval 0 to  $T$ , namely  $P'_T - P'_0$ , is made up of the total probability  $\sum_{\alpha''} P''_0 P(\alpha'' \alpha')$  of the system jumping into the state  $\alpha'$  from some other state  $\alpha''$ , minus the total probability  $P'_0 \sum_{\alpha''} P(\alpha' \alpha'')$  of its jumping out of the state  $\alpha'$ , during this time interval. Thus the ordinary laws of probability apply, showing that there is no interference between the different transition processes. If we had not averaged over the initial phases, there would have been such interference.

The integrand in (20) is the representative in a certain representation of the perturbing energy at time  $t$ . This representation is one that is approximately fixed in the Heisenberg picture, since if we put  $V = 0$  it would then be completely fixed in the Heisenberg picture. Hence we can, without spoiling the order of accuracy of



our result, replace the integral in (20) by  $\left(\alpha' \left| \int_0^T V_t dt \right| \alpha''\right)$  and obtain an alternative expression for the transition probability

$$P(\alpha''\alpha') = \hbar^{-2} \left| \left(\alpha' \left| \int_0^T V_t dt \right| \alpha''\right) \right|^2. \quad (21)$$

This provides a simple physical meaning for the non-diagonal elements of the matrix representing a dynamical variable if this dynamical variable can be regarded as the time integral of a perturbing energy.

#### 48. Application to Radiation

In the preceding section a general theory of the perturbation of an atomic system was developed, in which the perturbing energy could vary with the time in an arbitrary way. A perturbation of this kind can be realized in practice by allowing incident electromagnetic radiation to fall on the system. Let us see what our result (19) or (20) reduces to in this case.

If we neglect the effects of the magnetic field of the incident radiation, and if we further assume that the wave-lengths of the harmonic components of this radiation are all large compared with the dimensions of the atomic system, then the perturbing energy is simply the scalar product

$$V = (\mathbf{D}, \mathbf{\mathcal{E}}), \quad (22)$$

where  $\mathbf{D}$  is the total electric displacement of the system and  $\mathbf{\mathcal{E}}$  is the electric force of the incident radiation. We suppose  $\mathbf{\mathcal{E}}$  to be a given function of the time. If we take for simplicity the case when the incident radiation is plane polarized with its electric vector in a certain direction and let  $D$  denote the Cartesian component of  $\mathbf{D}$  in this direction, the expression (22) for  $V$  reduces to the ordinary product

$$V = D\mathcal{E},$$

where  $\mathcal{E}$  is the magnitude of the vector  $\mathbf{\mathcal{E}}$ . The matrix elements of  $V$  are

$$(\alpha' | V | \alpha'') = (\alpha' | D | \alpha'') \mathcal{E},$$

since  $\mathcal{E}$  is a number. Now  $(\alpha' | D | \alpha'')$  varies with the time  $t$  according to the Heisenberg law

$$(\alpha' | D | \alpha'') = (\alpha' | D | \alpha'')_0 e^{i(H'_0 - H''_0)t/\hbar},$$

$(\alpha' | D | \alpha'')_0$  being constant, and hence our expression (20) for the



transition probability becomes

$$P(\alpha' \alpha'') = \hbar^{-2} |(\alpha' | D | \alpha'')|^2 \left| \int_0^T \mathcal{E}_t e^{i(H'_0 - H''_0)t/\hbar} dt \right|^2. \quad (23)$$

If the incident radiation during the time interval 0 to  $T$  is resolved into its Fourier components, the energy crossing unit area per unit frequency range about the frequency  $\nu$  will be, according to classical electrodynamics,

$$E_\nu = \frac{c}{2\pi} \left| \int_0^T \mathcal{E}_t e^{2\pi i \nu t} dt \right|^2. \quad (24)$$

Comparing this with (23), we obtain

$$P(\alpha' \alpha'') = 2\pi c^{-1} \hbar^{-2} |(\alpha' | D | \alpha'')|^2 E_\nu, \quad (25)$$

where

$$\nu = |H'_0 - H''_0|/\hbar. \quad (26)$$

From this result we see in the first place that the transition probability depends only on that Fourier component of the incident radiation whose frequency  $\nu$  is connected with the change of energy by (26). This gives us *Bohr's Frequency Condition* and shows how the ideas of Bohr's atomic theory, which was the forerunner of quantum mechanics, can be fitted in with quantum mechanics.

The present elementary theory does not tell us anything about the energy of the field of radiation. It would be reasonable to assume, though, that the energy absorbed or liberated by the atomic system in the transition process comes from or goes into the component of the radiation with frequency  $\nu$  given by (26). This assumption will be justified by the more complete theory of radiation given in Chapter XI. The result (25) is then to be interpreted as the probability of the system, if initially in the state of lower energy, absorbing radiation and being carried to the upper state, and if initially in the upper state, being *stimulated* by the incident radiation to emit and fall to the lower state. The present theory does not account for the experimental fact that the system, if in the upper state with no incident radiation, can emit spontaneously and fall to the lower state, but this also will be accounted for by the more complete theory of Chapter XI.

The existence of the phenomenon of stimulated emission was inferred by Einstein,<sup>†</sup> long before the discovery of quantum mechanics, from a consideration of thermodynamic equilibrium between atoms

<sup>†</sup> Einstein, *Phys. Zeits.*, 18 (1917), 121.



and a field of black-body radiation satisfying Planck's law. Einstein showed that the transition probability for stimulated emission must equal that for absorption between the same pair of states and deduced a relation connecting this transition probability with that for spontaneous emission, which relation is in agreement with the theory of Chapter XI.

The matrix element  $(\alpha'|D|\alpha'')$  in (25) plays the part of the amplitude of one of the Fourier components of  $D$  in the classical theory of a multiply-periodic system interacting with radiation. In fact it was the idea of replacing classical Fourier components by matrix elements which led Heisenberg to the discovery of quantum mechanics in 1925. Heisenberg assumed that the formulas describing the interaction with radiation of a system in the quantum theory can be obtained from the classical formulas by substituting for the Fourier components of the total electric displacement of the system the corresponding matrix elements. According to this assumption applied to spontaneous emission, a system having an electric moment  $\mathbf{D}$  will, when in the state  $\alpha'$ , spontaneously emit radiation of frequency  $\nu = (H' - H'')/h$ , where  $H''$  is an energy-level, less than  $H'$ , of some state  $\alpha''$ , at the rate

$$\frac{4}{3} \frac{(2\pi\nu)^4}{c^3} |(\alpha'|\mathbf{D}|\alpha'')|^2. \quad (27)$$

The distribution of this radiation over the different directions of emission and its state of polarization for each direction will be the same as that for a classical electric dipole of moment equal to the real part of  $(\alpha'|\mathbf{D}|\alpha'')$ . To interpret this rate of emission of radiant energy as a transition probability, we must divide it by the quantum of energy of this frequency, namely  $h\nu$ , and call it the probability per unit time of this quantum being spontaneously emitted, with the atomic system simultaneously dropping to the state  $\alpha''$  of lower energy. These assumptions of Heisenberg are justified by the present radiation theory, supplemented by the spontaneous transition theory of Chapter XI.

#### 49. Transitions caused by a Perturbation Independent of the Time

The perturbation method of § 47 is still valid when the perturbing energy  $V$  does not involve the time  $t$  explicitly. Since the total Hamiltonian  $H$  in this case does not involve  $t$  explicitly, we could now, if desired, deal with the system by the perturbation method of



§ 46 and find its stationary states. Whether this method would be convenient or not would depend on what we want to find out about the system. If what we have to calculate makes an explicit reference to the time, e.g. if we have to calculate the wave function at one time when we are given its value at another time, the method of § 47 would be the more convenient one.

Let us see what the result (20) for the transition probability becomes when  $V$  does not involve  $t$  explicitly. The matrix element  $(\alpha'|V|\alpha'')$  now varies with  $t$  according to the Heisenberg law and thus its time integral is

$$\begin{aligned}\int_0^T (\alpha'|V|\alpha'') dt &= (\alpha'|V|\alpha'')_0 \int_0^T e^{i(H'_0 - H''_0)t/\hbar} dt \\ &= (\alpha'|V|\alpha'')_0 \frac{e^{i(H'_0 - H''_0)T/\hbar} - 1}{i(H'_0 - H''_0)/\hbar},\end{aligned}$$

provided  $H'_0 \neq H''_0$ . Thus the transition probability (20) becomes

$$\begin{aligned}P(\alpha'\alpha'') &= |(\alpha'|V|\alpha'')|^2 [e^{i(H'_0 - H''_0)T/\hbar} - 1][e^{-i(H'_0 - H''_0)T/\hbar} - 1]/(H'_0 - H''_0)^2 \\ &= 2|(\alpha'|V|\alpha'')|^2 [1 - \cos\{(H'_0 - H''_0)T/\hbar\}]/(H'_0 - H''_0)^2.\end{aligned}\quad (28)$$

If  $H''_0$  differs appreciably from  $H'_0$  this transition probability is small and remains so for all values of  $T$ . This result is required by the law of the conservation of energy. The total energy  $H$  is constant and hence the proper-energy  $H_0$  (i.e. the energy with neglect of the part  $V$  due to the perturbation), being approximately equal to  $H$ , must be approximately constant. This means that if  $H_0$  initially has the numerical value  $H'_0$ , at any later time there must be only a small probability of its having a numerical value differing considerably from  $H'_0$ .

On the other hand, when the initial state  $\alpha'$  is such that there exists another state  $\alpha''$  having the same or very nearly the same proper-energy  $H_0$ , the probability of a transition to the final state  $\alpha''$  may be quite large. The case of physical interest now is that in which there is a continuous range of final states  $\alpha''$  having a continuous range of proper-energy levels  $H''_0$  passing through the value  $H'_0$  of the proper-energy of the initial state. The initial state must not be one of the continuous range of final states, but may be either a separate discrete state or one of another continuous range of states. We shall now have, remembering the rules of § 22 for the interpretation of probability amplitudes with continuous ranges of states, that, with  $P(\alpha'\alpha'')$



having the value (28), the probability of a transition to a final state within the small range  $\alpha''$  to  $\alpha'' + d\alpha''$  will be  $P(\alpha'\alpha'') d\alpha''$  when the initial state  $\alpha'$  is discrete and will be proportional to this quantity when  $\alpha'$  is one of a continuous range.

We may suppose that the  $\alpha$ 's describing the final state consist of  $H_0$  itself together with a number of other dynamical variables  $\beta$ , so that we have a representation like that of § 46 for the degenerate case. (The  $\beta$ 's, however, need have no meaning for the initial state  $\alpha'$ .) We shall suppose for definiteness that the  $\beta$ 's have only discrete eigenvalues. The total probability of a transition to a final state  $\alpha''$  for which the  $\beta$ 's have the values  $\beta''$  and  $H_0$  has any value (there will be a strong probability of its having a value near the initial value  $H'_0$ ) will now be (or be proportional to)

$$\begin{aligned} & \int P(\alpha'\alpha'') dH''_0 \\ &= 2 \int_{-\infty}^{\infty} |(\alpha'|V|H''_0\beta'')|^2 [1 - \cos\{(H'_0 - H''_0)T/\hbar\}]/(H'_0 - H''_0)^2 \cdot dH''_0 \quad (29) \\ &= 2T\hbar^{-1} \int_{-\infty}^{\infty} |(\alpha'|V|H'_0 + \hbar x/T, \beta'')|^2 [1 - \cos x]/x^2 dx \end{aligned}$$

if one makes the substitution  $(H''_0 - H'_0)T/\hbar = x$ . For large values of  $T$  this reduces to

$$\begin{aligned} 2T\hbar^{-1} |(\alpha'|V|H'_0\beta'')|^2 \int_{-\infty}^{\infty} [1 - \cos x]/x^2 dx \\ = 2\pi T\hbar^{-1} |(\alpha'|V|H'_0\beta'')|^2. \end{aligned} \quad (30)$$

Thus the total probability up to time  $T$  of a transition to a final state for which the  $\beta$ 's have the values  $\beta''$  is proportional to  $T$ . There is therefore a definite probability coefficient, or probability per unit time, for the transition process under consideration, having the value

$$2\pi\hbar^{-1} |(\alpha'|V|H'_0\beta'')|^2. \quad (31)$$

It is proportional to the square of the modulus of the matrix element, associated with this transition, of the perturbing energy.

In order that the approximations used in deriving (30) may be valid, the time  $T$  must be not too small and not too large. It must be large compared with the periods of the atomic system in order that the evaluation of the integral (29) leading to the result (30) may be correct, while it must not be excessively large or else the general formula (20) will break down. In fact one could make the probability



(30) greater than unity by taking  $T$  large enough. The upper limit to  $T$  is fixed by the condition that the probability (20) or (30) must be small compared with unity. There is no difficulty in  $T$  satisfying both these conditions simultaneously provided the perturbing energy  $V$  is sufficiently small.

## 50. The Anomalous Zeeman Effect

One of the simplest examples of the perturbation method of § 46 is the calculation of the first-order change in the energy-levels of a general atom caused by a uniform magnetic field. The problem of a hydrogen atom in a uniform magnetic field has already been dealt with in § 43 and was so simple that perturbation theory was unnecessary. The case of a general atom is not much more complicated when we make a few approximations such that we can set up a simple model for the atom.

We first of all consider the atom in the absence of the magnetic field along the lines indicated in § 44 and look for angular momenta that are constants of the motion. The total angular momentum of the atom, the vector  $\mathbf{j}$  say, is certainly a constant of the motion. This angular momentum may be regarded as the sum of two parts, the total orbital angular momentum of all the electrons,  $\mathbf{l}$  say, and the total spin angular momentum,  $\mathbf{s}$  say. Thus we have  $\mathbf{j} = \mathbf{l} + \mathbf{s}$ . Now the effect of the spin magnetic moments on the motion of the electrons is small compared with the effect of the Coulomb forces and may be neglected as a first approximation. With this approximation the spin angular momentum of each electron is a constant of the motion, there being no forces tending to change its orientation. Thus  $\mathbf{s}$ , and hence also  $\mathbf{l}$ , will be constants of the motion. We now have the three constant angular momenta  $\mathbf{l}$ ,  $\mathbf{s}$ , and  $\mathbf{j}$ , related in the same way as the  $\mathbf{m}$ ,  $\boldsymbol{\mu}$ , and  $\mathbf{M}$  of § 44. The magnitudes,  $l$ ,  $s$ , and  $j$  say, of these angular momenta will be given by

$$\begin{aligned} l + \frac{1}{2}\hbar &= (l_x^2 + l_y^2 + l_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}} \\ s + \frac{1}{2}\hbar &= (s_x^2 + s_y^2 + s_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}} \\ j + \frac{1}{2}\hbar &= (j_x^2 + j_y^2 + j_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}}, \end{aligned}$$

corresponding to equation (12) of Chapter VII, and from (36) of that chapter we see that with given numerical values for  $l$  and  $s$  the possible numerical values for  $j$  are

$$l+s, \quad l+s-\hbar, \quad \dots, \quad |l-s|.$$



Let us consider a stationary state for which  $l$ ,  $s$ , and  $j$  have definite numerical values in agreement with the above scheme. The energy of this state will depend on  $l$ , but one might think that with neglect of the spin magnetic moments it would be independent of  $s$ , and also of the direction of the vector  $\mathbf{s}$  relative to  $\mathbf{l}$ , and thus of  $j$ . It will be found in Chapter X, however, that the energy depends very much on the magnitude  $s$  of the vector  $\mathbf{s}$ , although independent of its direction when one neglects the spin magnetic moments, on account of certain phenomena arising from the fact that the electrons are indistinguishable one from another. There are thus different energy-levels of the system for each different value of  $l$  and  $s$ . This means that  $l$  and  $s$  are functions of the energy, according to the general definition of a function given in § 11, since the  $l$  and  $s$  of a stationary state are fixed when the energy of that state is fixed.

We can now take into account the effect of the spin magnetic moments, treating it as a small perturbation according to the method of § 46. The energy of the unperturbed system will still be approximately a constant of the motion and hence  $l$  and  $s$ , being functions of this energy, will still be approximately constants of the motion. The directions of the vectors  $\mathbf{l}$  and  $\mathbf{s}$ , however, not being functions of the unperturbed energy, need not now be approximately constants of the motion and may undergo large secular variations. Since the vector  $\mathbf{j}$  is constant, the only possible variation of  $\mathbf{l}$  and  $\mathbf{s}$  is a precession about the vector  $\mathbf{j}$ . We thus have an approximate model of the atom consisting of the two vectors  $\mathbf{l}$  and  $\mathbf{s}$  of constant lengths precessing about their sum  $\mathbf{j}$ , which is a fixed vector. The energy is determined mainly by the magnitudes of  $\mathbf{l}$  and  $\mathbf{s}$  and depends only slightly on their relative directions, specified by  $j$ . Thus states with the same  $l$  and  $s$  and different  $j$  will have only slightly different energy-levels, forming what is called a *multiplet* term.

Let us now take this atomic model as our unperturbed system and suppose it to be subjected to a uniform magnetic field of magnitude  $\mathcal{H}$  in the direction of the  $z$ -axis. The extra energy due to this magnetic field will consist of a term

$$e\mathcal{H}/2mc \cdot (m_z + \hbar\sigma_z), \quad (32)$$

like the last term in equation (34) of Chapter VII, contributed by each electron, and will thus be altogether

$$e\mathcal{H}/2mc \cdot \sum (m_z + \hbar\sigma_z) = e\mathcal{H}/2mc \cdot (l_z + 2s_z) = e\mathcal{H}/2mc \cdot (j_z + s_z). \quad (33)$$



This is our perturbing energy  $V$ . We shall now use the method of § 46 to determine the changes in the energy-levels caused by this  $V$ . The method will be legitimate only provided the field is so weak that  $V$  is small compared with the energy differences within a multiplet.

Our unperturbed system is degenerate, on account of the direction of the vector  $\mathbf{j}$  being undetermined. We must therefore take, from the representative of  $V$  in a Heisenberg representation for the unperturbed system, those matrix elements that refer to one particular energy-level for their row and column, and obtain the eigenvalues of the matrix thus formed. We can do this best by first splitting up  $V$  into two parts, one of which is a constant of the unperturbed motion, so that its representative contains only matrix elements referring to the same unperturbed energy-level for their row and column, while the representative of the other contains only matrix elements referring to two different unperturbed energy-levels for their row and column, so that this second part does not affect the first-order perturbation. The term involving  $j_z$  in (33) is a constant of the unperturbed motion and thus belongs entirely to the first part. For the term involving  $s_z$  we have

$$s_z(j_x^2 + j_y^2 + j_z^2) = j_z(s_x j_x + s_y j_y + s_z j_z) + (s_z j_x - j_z s_x) j_x + (s_z j_y - j_z s_y) j_y$$

or

$$s_z = \frac{j_z}{j(j+\hbar)} \frac{1}{2} [j(j+\hbar) - l(l+\hbar) + s(s+\hbar)] - \frac{1}{j(j+\hbar)} [\gamma_y j_x - \gamma_x j_y], \quad (34)$$

where

$$\left. \begin{aligned} \gamma_x &= s_z j_y - j_z s_y = s_z l_y - l_z s_y = l_y s_z - l_z s_y \\ \gamma_y &= j_z s_x - s_z j_x = l_z s_x - s_z l_x = l_z s_x - l_x s_z \end{aligned} \right\} \quad (35)$$

The first term in this expression for  $s_z$  is a constant of the unperturbed motion and thus belongs entirely to the first part, while the second term, as we shall now see, belongs entirely to the second part.

Corresponding to (35) we can introduce

$$\gamma_z = l_x s_y - l_y s_x.$$

It can now easily be verified that

$$j_x \gamma_x + j_y \gamma_y + j_z \gamma_z = 0$$

and that

$$[j_z, \gamma_x] = \gamma_y, \quad [j_z, \gamma_y] = -\gamma_x, \quad [j_z, \gamma_z] = 0.$$

These relations are of the same form as the relations (3), (4), and (5) of Chapter VII, so that our  $\gamma_x, \gamma_y, \gamma_z$  are connected with the angular momentum  $\mathbf{j}$  in the same way in which the  $x, y, z$  of Chapter VII were connected with the angular momentum  $\mathbf{m}$ . We can thus take



over the analysis of § 42, in which the condition was obtained for the non-vanishing of a matrix element of  $x$ ,  $y$ , or  $z$  in a representation in which  $k$  is diagonal. We find in this way that the only non-vanishing matrix elements of  $\gamma_x$ ,  $\gamma_y$  and  $\gamma_z$  in a representation in which  $j$  is diagonal are those referring to transitions in which  $j$  changes by  $\pm\hbar$ . The coefficients of  $\gamma_x$  and  $\gamma_y$  in the second term on the right-hand side of (34) commute with  $j$ , so that the representative of the whole of this term will contain only matrix elements referring to transitions in which  $j$  changes by  $\pm\hbar$ , and thus referring to two different energy-levels of the unperturbed system.

Hence the perturbing energy  $V$  becomes, when we neglect that part of it whose representative consists of matrix elements referring to two different unperturbed energy-levels,

$$\frac{e\mathcal{H}}{2mc} j_z \left\{ 1 + \frac{1}{2} \frac{j(j+\hbar) - l(l+\hbar) + s(s+\hbar)}{j(j+\hbar)} \right\}. \quad (36)$$

The eigenvalues of this give the first-order changes in the energy-levels. We can make the representative of this expression diagonal by choosing our representation such that  $j_z$  is diagonal, i.e. by taking the basic states to be spacially quantized in the  $z$ -direction. The expression (36) then gives us directly the first-order changes in the energy-levels caused by the magnetic field. This expression is known as Landé's formula.

The result (36) holds only provided the perturbing energy  $V$  is small compared with the energy differences within a multiplet. For larger values of  $V$  a more complicated theory is required. For very strong fields, however, for which  $V$  is large compared with the energy differences within a multiplet, the theory is again very simple. We may now neglect altogether the energy of the spin magnetic moments for the atom with no external field, so that for our unperturbed system the vectors  $\mathbf{l}$  and  $\mathbf{s}$  themselves are constants of the motion, and not merely their magnitudes  $l$  and  $s$ . Our perturbing energy  $V$ , which is still  $e\mathcal{H}/2mc \cdot (j_z + s_z)$ , is now a constant of the motion for the unperturbed system, so that its eigenvalues give directly the changes in the energy-levels. These eigenvalues are integral or half-odd integral multiples of  $e\mathcal{H}\hbar/2mc$  according to whether the number of electrons in the atom is even or odd.



## IX

### COLLISION PROBLEMS

#### 51. General Remarks

IN this chapter we shall investigate problems connected with a particle which, coming from infinity, encounters or 'collides with' some atomic system and, after being scattered through a certain angle, goes off to infinity again. The atomic system which does the scattering we shall call, for brevity, the *scatterer*. We thus have a dynamical system composed of an incident particle and a scatterer interacting with each other, which we must deal with according to the laws of quantum mechanics, and for which we must, in particular, calculate the probability of scattering through any given angle. This problem was first solved by Born by a method substantially equivalent to that of the next section. We must take into account the possibility that the scatterer, considered as a system by itself, may have a number of different stationary states and that if it is initially in one of these states when the particle arrives from infinity, it may be left in a different one when the particle goes off to infinity again. The colliding particle may thus induce transitions in the scatterer.

The Hamiltonian for the whole system of scatterer plus particle will not involve the time explicitly, so that this whole system will have stationary states represented by simply-periodic solutions of Schrödinger's wave equation. The meaning of these stationary states requires a little care to be properly understood. It is evident that for *any* state of motion of the system the particle will spend nearly all its time at infinity, so that the time average of the probability of the particle being in any finite volume will be zero. Now for a *stationary* state the probability of the particle being in a given finite volume, like any other result of observation, must be independent of the time, and hence this probability will equal its time average, which we have seen is zero. We shall thus be interested only in the relative probabilities of the particle being in different finite volumes, their absolute values being all zero. Mathematically we have that if the  $\psi$  representing a stationary state is normalized correctly for physical interpretation, i.e. such that  $\phi\psi = 1$ , and if we let  $Q$  denote that observable (which is a certain function of the position of the particle) that is equal to unity if the particle is in a given finite volume and zero



otherwise, then  $\phi Q\psi = 0$ , meaning that the average value of  $Q$ , i.e. the probability of the particle being in the given volume, is zero. It would therefore be more convenient for us to denote the stationary state by a  $\psi$  normalized to infinity, i.e. for which  $\phi\psi = \infty$ , the infinity being such as to make  $\phi Q\psi$  finite. This finite  $\phi Q\psi$  would then give the relative probability of the particle being in the given volume.

In picturing a state of a system represented by a  $\psi$  which is not normalized correctly for physical interpretation, but for which  $\phi\psi = n$  say, it may be convenient to suppose that we have  $n$  similar systems all occupying the same space but with no interaction between them, so that each one follows out its own motion independently of the others, as we had in the theory of the Gibbs ensemble in § 37. We can then interpret  $\phi\alpha\psi$ , where  $\alpha$  is any observable, directly as the total  $\alpha$  for all the  $n$  systems. In applying these ideas to the above-mentioned  $\psi$  normalized to infinity, representing a stationary state of the system of scatterer plus colliding particle, we should picture an infinite number of such systems with the scatterers all located at the same point and the particles distributed continuously throughout space. The number of particles in a given finite volume would be pictured as  $\phi Q\psi$ ,  $Q$  being the observable defined above, which has the value unity when the particle is in the given volume and zero otherwise. If the  $\psi$  is represented by a Schrödinger wave function involving the Cartesian coordinates of the particle, then the square of the modulus of the wave function could be interpreted directly as the density of particles in the picture. One must remember, however, that *each of these particles has its own individual scatterer*. Different particles may belong to scatterers in different states. There will thus be one particle density for each state of the scatterer, namely, the density of those particles belonging to scatterers in that state. This is taken account of by the wave function involving variables describing the state of the scatterer in addition to those describing the position of the particle.

For determining scattering coefficients we have to investigate *stationary states* of the whole system of scatterer plus particle. For instance, if we want to determine the probability of scattering in various directions when the scatterer is initially in a given stationary state and the incident particle has initially a given velocity in a given direction, we must investigate that stationary state of the whole system whose picture, according to the above method, contains at



great distances from the point of location of the scatterers only particles moving with the given initial velocity and direction and belonging each to a scatterer in the given initial stationary state, together with particles moving *outward* from the point of location of the scatterers and belonging possibly to scatterers in various stationary states. This picture corresponds closely to the actual state of affairs in an experimental determination of scattering coefficients, with the difference that the picture really describes only one *actual* system of scatterer plus particle. The distribution of outward moving particles at infinity in the picture gives us immediately all the information about scattering coefficients that could be obtained by experiment. For practical calculations about the stationary state described by this picture one may use the perturbation method of § 46, taking as unperturbed system, for example, that for which there is no interaction between the scatterer and particle.

In dealing with collision problems, a further possibility to be taken into consideration is that the scatterer may perhaps be capable of absorbing and re-emitting the particle. This possibility arises when there exists one or more *states of absorption* of the whole system, a state of absorption being an *approximately* stationary state which is closed in the sense of § 40 (i.e. for which the probability of the particle being at a greater distance than  $r$  from the scatterer tends to zero as  $r \rightarrow \infty$ ). Since a state of absorption is only approximately stationary, its property of being closed will be only a transient one, and after a sufficient lapse of time there will be a finite probability of the particle being on its way to infinity. Physically this means there is a finite probability of spontaneous emission of the particle. The fact that we had to use the word 'approximately' in stating the conditions required for the phenomena of emission and absorption to be able to occur shows that these conditions are not expressible in exact mathematical language. One can give a meaning to these phenomena only with reference to a perturbation method. They occur when the unperturbed system (of scatterer plus particle) has stationary states that are closed. The introduction of the perturbation now spoils the stationary property of these states and gives rise to spontaneous emission and its converse absorption.

For calculating absorption and emission probabilities it is necessary to deal with *non-stationary states* of the system, in contradistinction to the case for scattering coefficients, so that the perturbation method



of § 47 must be used. Thus for calculating an emission coefficient we must consider the non-stationary states of absorption described above. Again, since an absorption is always followed by a re-emission, it cannot be distinguished from a scattering in any experiment involving a steady state of affairs, corresponding to a stationary state of the system. The distinction can be made only by reference to a non-steady state of affairs, e.g. by use of a stream of incident particles that has a sharp beginning, so that the scattered particles will appear immediately after the incident particles meet the scatterers, while those that have been absorbed and re-emitted will begin to appear only some time later. This stream of particles would then be the picture of a certain non-stationary  $\psi$ , normalized to infinity, which could be used for calculating the absorption coefficient.

## 52. The Scattering Coefficient

We shall now consider the calculation of scattering coefficients, taking first the case when there is no absorption and emission, which means that our unperturbed system has no closed stationary states. We may conveniently take this unperturbed system to be that for which there is no interaction between the scatterer and particle. Its Hamiltonian will thus be of the form

$$H_0 = H_s + W, \quad (1)$$

where  $H_s$  is that for the scatterer alone and  $W$  that for the particle alone, namely, with neglect of relativistic variation of mass with velocity,

$$W = 1/2m \cdot (p_x^2 + p_y^2 + p_z^2). \quad (2)$$

The perturbing energy  $V$ , assumed small, will now be a function of the Cartesian coordinates of the particle  $x, y, z$ , and also, perhaps, of its momenta  $p_x, p_y, p_z$ , together with dynamical variables describing the scatterer.

Since we are now interested only in *stationary* states of the whole system, we can use the perturbation method of § 46. Our unperturbed system now necessarily has a continuous range of energy-levels, since it contains a free particle, and this gives rise to certain modifications in the perturbation method. The question of the change in the energy-levels caused by the perturbation, which was the main question of § 46, no longer has a meaning, and the convention in § 46 of using the same number of primes to denote nearly equal eigenvalues of  $H_0$  and  $H$  now drops out. Again, the splitting of energy-



levels which we had in § 46 when the unperturbed system is degenerate cannot now arise, since if the unperturbed system is degenerate the perturbed one, which must also have a continuous range of energy-levels, will also be degenerate to exactly the same extent.

We again use the general scheme of equations developed at the beginning of § 46, equations (1) to (4) there, but we now take our unperturbed stationary state forming the zero-order approximation to belong to an energy-level  $H'_0$  just equal to the energy-level  $H'$  of our perturbed stationary state. We put  $H'_0 = H' = E$  say. Thus the  $a$ 's introduced in the second of equations (3) § 46 are now all zero and the second of equations (4) there now reads

$$\{E - H_0\}\psi_1 = V\psi_0. \quad (3)$$

Similarly, the third of equations (4) § 46 now reads

$$\{E - H_0\}\psi_2 = V\psi_1. \quad (4)$$

We shall proceed to solve equation (3) and to obtain the scattering coefficient to the first order. We shall need equation (4) later.

Let  $\alpha$  denote a complete set of commuting observables describing the scatterer, which are constants of the motion when the scatterer is alone and may thus be used for labelling the stationary states of the scatterer. This requires that  $H_s$  shall commute with the  $\alpha$ 's and be a function of them. We can now take a representation of the whole system in which the  $\alpha$ 's and  $x, y, z$ , the coordinates of the particle, are diagonal. This will make  $H_s$  diagonal. Let  $\psi_0$  be represented by  $(\mathbf{x}\alpha|0)$  and  $\psi_1$  by  $(\mathbf{x}\alpha|1)$ , the single variable  $\mathbf{x}$  being written in the wave function to denote  $x, y, z$ . In the same way the single differential  $d\mathbf{x}$  will be written to denote the product  $dx dy dz$ . Equation (3), written in terms of representatives, becomes, with the help of (1) and (2),

$$\{E - H_s(\alpha') + \hbar^2/2m \cdot \nabla^2\}(\mathbf{x}\alpha'|1) = \sum_{\alpha''} \int (\mathbf{x}\alpha'|V|\mathbf{x}''\alpha'') d\mathbf{x}'' (\mathbf{x}''\alpha''|0). \quad (5)$$

Suppose that the incident particle has the momentum  $\mathbf{p}^0$  and that the initial stationary state of the scatterer is  $\alpha^0$ . The stationary state  $\psi_0$  of our unperturbed system is now the one for which  $\mathbf{p} = \mathbf{p}^0$  and  $\alpha = \alpha^0$ , and hence its representative is of the form

$$(\mathbf{x}\alpha|0) = \delta_{\alpha\alpha^0} e^{i(\mathbf{p}^0, \mathbf{x})/\hbar}. \quad (6)$$

This makes equation (5) reduce to

$$\{E - H_s(\alpha') + \hbar^2/2m \cdot \nabla^2\}(\mathbf{x}\alpha'|1) = \int (\mathbf{x}\alpha'|V|\mathbf{x}^0\alpha^0) d\mathbf{x}^0 e^{i(\mathbf{p}^0, \mathbf{x}^0)/\hbar}$$



$$\text{or} \quad \{k^2 + \nabla^2\}(\mathbf{x}\alpha'|1) = F, \quad (7)$$

$$\text{where} \quad k^2 = 2m\hbar^{-2}\{E - H_s(\alpha')\} \quad (8)$$

$$\text{and} \quad F = 2m\hbar^{-2} \int (\mathbf{x}\alpha'|V|\mathbf{x}^0\alpha^0) d\mathbf{x}^0 e^{i(\mathbf{p}^0, \mathbf{x}^0)/\hbar}, \quad (9)$$

a definite function of  $x, y, z$ , and  $\alpha'$ . We must also have

$$E = H'_0 = H_s(\alpha^0) + \mathbf{p}^{02}/2m. \quad (10)$$

Our problem now is to obtain a solution  $(\mathbf{x}\alpha'|1)$  of (7) which, for values of  $x, y, z$  denoting points far from the scatterer, represents only outward moving particles. The square of its modulus,  $|(\mathbf{x}\alpha'|1)|^2$ , will then give the density of scattered particles belonging to scatterers in the state  $\alpha'$  when the density of the incident particles is  $|(\mathbf{x}\alpha^0|0)|^2$ , which is unity. If we transform to polar coordinates  $r, \theta, \phi$ , equation (7) becomes

$$\left\{k^2 + \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}\right\} (r\theta\phi\alpha'|1) = F. \quad (11)$$

Now  $F$  must tend to zero as  $r \rightarrow \infty$ , on account of the physical requirement that the interaction energy between the scatterer and particle must tend to zero as the distance between them tends to infinity. If we neglect  $F$  in (11) altogether, an approximate solution for large  $r$  is

$$(r\theta\phi\alpha'|1) = u(\theta\phi\alpha')r^{-1}e^{ikr}, \quad (12)$$

where  $u$  is an arbitrary function of  $\theta, \phi$  and  $\alpha'$ , since this expression substituted in the left-hand side of (11) gives a result of order  $r^{-3}$ . When we do not neglect  $F$ , the solution of (11) will still be of the form (12) for large  $r$ , provided  $F$  tends to zero sufficiently rapidly as  $r \rightarrow \infty$ , but the function  $u$  will now be definite and determined by the solution for smaller values of  $r$ .

For values  $\alpha'$  of the  $\alpha$ 's such that  $k^2$ , defined by (8), is positive, the  $k$  in (12) must be chosen to be the positive square root of  $k^2$ , in order that (12) may represent only outward moving particles, i.e. particles for which the radial component of momentum  $p_r$ , represented by  $-i\hbar\partial/\partial r$  when it operates to the right, has a positive value. We now have that the density of scattered particles belonging to scatterers in state  $\alpha'$ , equal to the square of the modulus of (12), falls off with increasing  $r$  according to the inverse square law, as is physically necessary, and their angular distribution is given by  $|u(\theta\phi\alpha')|^2$ . Further, the magnitude,  $P'$  say, of the momentum of these scattered particles must equal  $k\hbar$ , to make the exponential in (12) of the



form  $e^{iP'r/\hbar}$ , so that their energy is equal to

$$\frac{P'^2}{2m} = \frac{k^2\hbar^2}{2m} = E - H_s(\alpha') = H_s(\alpha^0) - H_s(\alpha') + \frac{\mathbf{p}^2}{2m},$$

with the help of (8) and (10). This is just the energy of an incident particle, namely  $\mathbf{p}^2/2m$ , reduced by the increase in energy of the scatterer, namely  $H_s(\alpha') - H_s(\alpha^0)$ , in agreement with the law of conservation of energy. For values  $\alpha'$  of the  $\alpha$ 's such that  $k^2$  is negative there are no scattered particles, the total initial energy being insufficient for the scatterer to be left in the state  $\alpha'$ .

We must now evaluate  $u(\theta\phi\alpha')$  for a set of values  $\alpha'$  for the  $\alpha$ 's such that  $k^2$  is positive, and obtain the angular distribution of the scattered particles belonging to scatterers in state  $\alpha'$ . It is sufficient to evaluate  $u$  for the direction  $\theta = 0$  of the pole of the polar coordinates, since this direction is arbitrary. We make use of Green's theorem, which states that for any two functions of position  $A$  and  $B$  the volume integral  $\int (A\nabla^2 B - B\nabla^2 A) d\mathbf{x}$  taken over any volume equals the surface integral  $\int (A\partial B/\partial n - B\partial A/\partial n) dS$  taken over the boundary of the volume,  $\partial/\partial n$  denoting differentiation along the normal to the surface. We take

$$A = e^{-ikr\cos\theta} \quad B = (r\theta\phi\alpha'|1)$$

and apply the theorem to a large sphere with the origin as centre. The volume integrand is thus

$$\begin{aligned} e^{-ikr\cos\theta}\nabla^2(r\theta\phi\alpha'|1) - (r\theta\phi\alpha'|1)\nabla^2 e^{-ikr\cos\theta} \\ = e^{-ikr\cos\theta}(\nabla^2 + k^2)(r\theta\phi\alpha'|1) = e^{-ikr\cos\theta}F \end{aligned}$$

from (7) or (11), while the surface integrand is, with the help of (12),

$$\begin{aligned} e^{-ikr\cos\theta}\frac{\partial}{\partial r}(r\theta\phi\alpha'|1) - (r\theta\phi\alpha'|1)\frac{\partial}{\partial r}e^{-ikr\cos\theta} \\ = e^{-ikr\cos\theta}u\left(-\frac{1}{r^2} + \frac{ik}{r}\right)e^{ikr} + i\frac{u}{r}e^{ikr}k\cos\theta e^{-ikr\cos\theta} \\ = ikur^{-1}(1 + \cos\theta)e^{ikr(1-\cos\theta)} \end{aligned}$$

with neglect of  $r^{-2}$ . Hence we get

$$\begin{aligned} \int e^{-ikr\cos\theta}F d\mathbf{x} &= \int_0^{2\pi} d\phi \int_0^\pi r^2 \sin\theta d\theta \cdot ikur^{-1}(1 + \cos\theta)e^{ikr(1-\cos\theta)} \\ &= ikr \int_0^{2\pi} d\phi \int_0^2 d\gamma \cdot u(\theta\phi\alpha')(2 - \gamma)e^{ikr\gamma}, \end{aligned}$$



where  $\gamma = 1 - \cos \theta$ , the volume integral on the left being taken over the whole of space. The right-hand side becomes, on being integrated by parts with respect to  $\gamma$ ,

$$\int_0^{2\pi} d\phi \left\{ [u(\theta\phi\alpha')(2-\gamma)e^{ikr\gamma}]_{\gamma=0}^{\gamma=2} - \int_0^2 d\gamma \cdot e^{ikr\gamma} \frac{\partial}{\partial \gamma} [u(\theta\phi\alpha')(2-\gamma)] \right\}.$$

The second term in the  $\{ \}$  brackets is of the order of magnitude of  $r^{-1}$ , as would be revealed by further partial integrations, and may therefore be neglected. We are thus left with

$$\int e^{-ikr \cos \theta} F \, d\mathbf{x} = -2 \int_0^{2\pi} d\phi u(0\phi\alpha') = -4\pi u(0\phi\alpha'),$$

giving the value of  $u(\theta\phi\alpha')$  for the direction  $\theta = 0$ .

This result may be written

$$u(0\phi\alpha') = -(4\pi)^{-1} \int e^{-iP'r \cos \theta / \hbar} F \, d\mathbf{x}, \quad (13)$$

since  $P' = k\hbar$ . If the vector  $\mathbf{p}'$  denotes the momentum of the scattered electrons coming off in a certain direction (and is thus of magnitude  $P'$ ), the value of  $u$  for this direction will be

$$u(\theta'\phi'\alpha') = -(4\pi)^{-1} \int e^{-i(\mathbf{p}', \mathbf{x})/\hbar} F \, d\mathbf{x},$$

as follows from (13) if one takes this direction to be the pole of the polar coordinates. This becomes, with the help of (9),

$$\begin{aligned} u(\theta'\phi'\alpha') &= -(2\pi)^{-1} m \hbar^{-2} \iint e^{-i(\mathbf{p}', \mathbf{x})/\hbar} d\mathbf{x} (\mathbf{x}\alpha' | V | \mathbf{x}^0\alpha^0) d\mathbf{x}^0 e^{i(\mathbf{p}^0, \mathbf{x}^0)/\hbar} \\ &= -2\pi m \hbar (\mathbf{p}'\alpha' | V | \mathbf{p}^0\alpha^0), \end{aligned} \quad (14)$$

when one makes a transformation from the coordinates  $\mathbf{x}$  to the momenta  $\mathbf{p}$  of the particle, using the transformation function (43) of Chapter V. The single letter  $\mathbf{p}$  is here used to denote the three components of momentum.

The density of scattered particles belonging to scatterers in state  $\alpha'$  is now given by  $|u(\theta'\phi'\alpha')|^2/r^2$ . Since their velocity is  $P'/m$ , the rate at which these particles appear per unit solid angle about the direction of the vector  $\mathbf{p}'$  will be  $P'/m \cdot |u(\theta'\phi'\alpha')|^2$ . The density of the incident particles is, as we have seen, unity, so that the number of incident particles crossing unit area per unit time is equal to their velocity  $P^0/m$ , where  $P^0$  is the magnitude of  $\mathbf{p}^0$ . Hence the effective area that must be hit by an incident particle in order to be scattered



in a unit solid angle about the direction  $\mathbf{p}'$  and then belong to a scatterer in state  $\alpha'$  will be

$$P'/P^0 \cdot |u(\theta' \phi' \alpha')|^2 = 4\pi^2 m^2 \hbar^2 P'/P^0 \cdot |(\mathbf{p}'\alpha'|V|\mathbf{p}^0\alpha^0)|^2. \quad (15)$$

This is the scattering coefficient for transitions  $\alpha^0 \rightarrow \alpha'$  of the scatterer. It depends on that matrix element  $(\mathbf{p}'\alpha'|V|\mathbf{p}^0\alpha^0)$  of the perturbing energy  $V$  whose column  $\mathbf{p}^0\alpha^0$  and whose row  $\mathbf{p}'\alpha'$  refer respectively to the initial and final states of the unperturbed system, between which the scattering transition process takes place. The result (15) is thus in some ways analogous to the result (20) or (21) of Chapter VIII, although the numerical coefficients are different in the two cases, corresponding to the different natures of the two transition processes.

### 53. Solution with the $\mathbf{p}$ -Representation

The result (15) for the scattering coefficient makes a reference only to that representation in which the momentum  $\mathbf{p}$  is diagonal. One would thus expect to be able to get a more direct proof of the result by working all the time in the  $\mathbf{p}$ -representation, instead of working in the  $\mathbf{x}$ -representation and transforming at the end to the  $\mathbf{p}$ -representation, as was done in § 52. This would not at first sight appear to be a great improvement, as the lack of directness of the  $\mathbf{x}$ -representation method is offset by its greater 'Anschaulichkeit', it being possible to picture the square of the modulus of the  $\mathbf{x}$ -representative of a state as the density of a stream of particles in process of being scattered. The  $\mathbf{x}$ -representation method has, however, other more serious disadvantages. One of the main applications of the theory of collisions is to the case of photons as incident particles. Now a photon is not a simple particle but has a polarization. It is evident from classical electromagnetic theory that a photon with a definite momentum, i.e. one moving in a definite direction with a definite frequency, may have a definite state of polarization (linear, circular, etc.), while a photon with a definite position, which is to be pictured as an electromagnetic disturbance confined to a very small volume, cannot have any definite polarization. These facts mean that the polarization observable of a photon commutes with its momentum but not with its position. This results in the  $\mathbf{p}$ -representation method being immediately applicable to the case of photons, it being only necessary to introduce the polarizing variable into the representatives and treat it along with the  $\alpha$ 's describing the scatterer, while the  $\mathbf{x}$ -representation method is not applicable. Further, in dealing with



photons, it is necessary to take the relativistic variation of mass with velocity into account. This can easily be done in the  $\mathbf{p}$ -representation method, but not so easily in the  $\mathbf{x}$ -representation method.

Equation (3) still holds when the relativistic variation of mass with velocity is taken into account for the particle, but  $W$  is now given by

$$W^2/c^2 = m^2c^2 + P^2 = m^2c^2 + p_x^2 + p_y^2 + p_z^2 \quad (16)$$

instead of by (2). Written in terms of  $\mathbf{p}$ -representatives, equation (3) becomes

$$\{E - H_s(\alpha') - W\}(\mathbf{p}\alpha'|1) = \sum_{\alpha''} \int (\mathbf{p}\alpha'|V|\mathbf{p}''\alpha'') d\mathbf{p}'' (\mathbf{p}''\alpha''|0),$$

$W$  being here understood as a definite function of  $p_x, p_y, p_z$  given by (16). This may be written

$$\{W' - W\}(\mathbf{p}\alpha'|1) = \sum_{\alpha''} \int (\mathbf{p}\alpha'|V|\mathbf{p}''\alpha'') d\mathbf{p}'' (\mathbf{p}''\alpha''|0), \quad (17)$$

$$\text{where} \quad W' = E - H_s(\alpha') \quad (18)$$

and is the energy required by the law of conservation of energy for a scattered particle belonging to a scatterer in state  $\alpha'$ . The  $\mathbf{p}$ -representative of  $\psi_0$ , obtained by transforming (6) with the transformation function (43) of Chapter V, is

$$(\mathbf{p}\alpha|0) = h^3 \delta_{\alpha\alpha^0} \delta(\mathbf{p} - \mathbf{p}^0), \quad (19)$$

as may be verified most easily by transforming this back to the  $\mathbf{x}$ -representation. The  $\delta(\mathbf{p} - \mathbf{p}^0)$  means the product

$$\delta(p_x - p_x^0) \delta(p_y - p_y^0) \delta(p_z - p_z^0).$$

Equation (17) now becomes

$$\{W' - W\}(\mathbf{p}\alpha'|1) = h^3 (\mathbf{p}\alpha'|V|\mathbf{p}^0\alpha^0). \quad (20)$$

We now make a canonical transformation from the Cartesian coordinates  $p_x, p_y, p_z$  of  $\mathbf{p}$  to its polar coordinates  $P, \omega, \chi$ , given by

$$p_x = P \cos \omega \quad p_y = P \sin \omega \cos \chi \quad p_z = P \sin \omega \sin \chi.$$

If in the new representation we take the weight function  $P^2 \sin \omega$ , then the weight attached to any volume of  $\mathbf{p}$ -space will be the same as in the previous  $\mathbf{p}$ -representation, so that the canonical transformation will mean simply a relabelling of the rows and columns of the matrices without any alteration of the matrix elements or of the coordinates representing a state. Thus (20) will become in the new representation

$$\{W' - W\}(P\omega\chi\alpha'|1) = h^3 (P\omega\chi\alpha'|V|P^0\omega^0\chi^0\alpha^0), \quad (21)$$

$W$  being now a function of the single variable  $P$ .



The coefficient of  $(P\omega\chi\alpha'|1)$ , namely  $\{W'-W\}$ , is now simply a multiplying factor and not a differential operator as it was with the  $x$ -representation method. We can therefore divide out by this factor and obtain an explicit expression for  $(P\omega\chi\alpha'|1)$ . When, however,  $\alpha'$  is such that  $W'$ , defined by (18), is greater than  $mc^2$ , this factor will have the value zero for a certain point in the domain of the variable  $P$ , namely the point  $P = P'$ , given in terms of  $W'$  by (16). The function  $(P\omega\chi\alpha'|1)$  will then have a singularity at this point. This singularity shows that  $(P\omega\chi\alpha'|1)$  represents an infinite number of particles moving about at great distances from the scatterers with energies indefinitely close to  $W'$  and it is therefore this singularity that we have to study to get the angular distribution of the particles at infinity.

The result of dividing out (21) by the factor  $\{W'-W\}$  is

$$(P\omega\chi\alpha'|1) = h^3(P\omega\chi\alpha'|V|P^0\omega^0\chi^0\alpha^0)/\{W'-W\} + \lambda(\omega\chi\alpha')\delta(W'-W), \quad \dots (22)$$

where  $\lambda$  is an arbitrary function of  $\omega$ ,  $\chi$  and  $\alpha'$ , since when an arbitrary multiple of  $\delta(W'-W)$  is multiplied by  $W'-W$  the product will vanish. To give a meaning to the first term on the right-hand side of (22), we make the convention that its integral with respect to  $P$  over a range that includes the value  $P'$  is the limit when  $\epsilon \rightarrow 0$  of the integral when the small domain  $P'-\epsilon$  to  $P'+\epsilon$  is excluded from the range of integration. This is sufficient to make the meaning of (22) precise, since we are interested effectively only in the integrals of the representatives of states when the representation has continuous ranges of rows and columns. We see that equation (21) is inadequate to determine the representative  $(P\omega\chi\alpha'|1)$  completely, on account of the arbitrary function  $\lambda$  occurring in (22). We must choose this  $\lambda$  such that  $(P\omega\chi\alpha'|1)$  represents only outward moving particles, since we want the only inward moving particles to be those represented by (19).

Let us take first the general case when the representative  $(P\omega\chi|)$  of a state of the particle satisfies an equation of the type

$$\{W'-W\}(P\omega\chi|) = f(P\omega\chi), \quad (23)$$

where  $f(P\omega\chi)$  is any function of  $P$ ,  $\omega$  and  $\chi$ , and  $W'$  is a number greater than  $mc^2$ , so that  $(P\omega\chi|)$  is of the form

$$(P\omega\chi|) = f(P\omega\chi)/\{W'-W\} + \lambda(\omega\chi)\delta(W'-W), \quad (24)$$

and let us determine now what  $\lambda$  must be in order that  $(P\omega\chi|)$  may



represent only outward moving particles. We can do this by transforming  $(P\omega\chi|)$  to the  $\mathbf{x}$ -representation, or rather the  $(r\theta\phi)$ -representation, and comparing it with (12) for large values of  $r$ . The transformation function is

$$(r\theta\phi|P\omega\chi) = h^{-\frac{1}{2}} e^{i(\mathbf{p}, \mathbf{x})/\hbar} = h^{-\frac{1}{2}} e^{iPr[\cos\omega\cos\theta + \sin\omega\sin\theta\cos(\chi-\phi)]/\hbar}.$$

For the direction  $\theta = 0$  we find

$$\begin{aligned} (r0\phi|) &= h^{-\frac{1}{2}} \int_0^\infty P^2 dP \int_0^{2\pi} d\chi \int_0^\pi \sin\omega d\omega e^{iPr\cos\omega/\hbar} (P\omega\chi|) \\ &= h^{-\frac{1}{2}} \int_0^\infty P^2 dP \int_0^{2\pi} d\chi \left\{ - \left[ \frac{e^{iPr\cos\omega/\hbar}}{iPr/\hbar} (P\omega\chi|) \right]_{\omega=0}^{\omega=\pi} + \right. \\ &\quad \left. + \int_0^\pi d\omega \frac{e^{iPr\cos\omega/\hbar}}{iPr/\hbar} \frac{\partial}{\partial\omega} (P\omega\chi|) \right\}. \end{aligned}$$

The second term in the  $\{ \}$  brackets is of order  $r^{-2}$ , as may be verified by further partial integrations with respect to  $\omega$ , and can therefore be neglected. We are left with

$$\begin{aligned} (r0\phi|) &= ih^{-\frac{1}{2}} (2\pi r)^{-1} \int_0^\infty P dP \int_0^{2\pi} d\chi \{ e^{-iPr/\hbar} (P\pi\chi|) - e^{iPr/\hbar} (P0\chi|) \} \\ &= ih^{-\frac{1}{2}} r^{-1} \int_0^\infty P dP \{ e^{-iPr/\hbar} (P\pi\chi|) - e^{iPr/\hbar} (P0\chi|) \}. \end{aligned} \quad (25)$$

When we substitute for  $(P\omega\chi|)$  its value given by (24), the first term in the integrand in (25) gives

$$ih^{-\frac{1}{2}} r^{-1} \int_0^\infty P dP e^{-iPr/\hbar} \{ f(P\pi\chi)/(W' - W) + \lambda(\pi\chi)\delta(W' - W) \}. \quad (26)$$

The term involving  $\delta(W' - W)$  here may be integrated immediately and gives, when one uses the relation  $P dP = W dW/c^2$ , which follows from (16),

$$\begin{aligned} ih^{-\frac{1}{2}} c^{-2} r^{-1} \int_{mc^2}^\infty W dW e^{-iPr/\hbar} \lambda(\pi\chi) \delta(W' - W) \\ = ih^{-\frac{1}{2}} c^{-2} r^{-1} W' \lambda(\pi\chi) e^{-iP'r/\hbar}. \end{aligned} \quad (27)$$

To integrate the other term in (26) we use the formula that

$$\int_0^\infty g(P) \frac{e^{-iPr/\hbar}}{P' - P} dP = g(P') \int_0^\infty \frac{e^{-iPr/\hbar}}{P' - P} dP, \quad (28)$$



with neglect of terms involving  $r^{-1}$ , for any continuous function  $g(P)$ , which formula holds since  $\int_0^\infty K(P)e^{-iPr/\hbar} dP$  is of order  $r^{-1}$  for any continuous function  $K(P)$  and since the difference

$$g(P)/(P'-P) - g(P')/(P'-P)$$

is continuous. The right-hand side of (28), when evaluated with neglect of terms involving  $r^{-1}$ , and also with neglect of the small domain  $P'-\epsilon$  to  $P'+\epsilon$  in the domain of integration, gives

$$\begin{aligned} g(P') \int_{-\infty}^{\infty} \frac{e^{-iPr/\hbar}}{P'-P} dP &= g(P')e^{-iP'r/\hbar} \int_{-\infty}^{\infty} \frac{e^{i(P'-P)r/\hbar}}{P'-P} dP \\ &= ig(P')e^{-iP'r/\hbar} \int_{-\infty}^{\infty} \frac{\sin(P'-P)r/\hbar}{P'-P} dP = i\pi g(P')e^{-iP'r/\hbar}. \end{aligned} \quad (29)$$

In our present example  $g(P)$  is

$$g(P) = ih^{-\frac{1}{2}}r^{-1}Pf(P\pi\chi)(P'-P)/(W'-W),$$

which has the limiting value when  $P = P'$ ,

$$g(P') = ih^{-\frac{1}{2}}r^{-1}P'f(P'\pi\chi)W'/P'^2 = ih^{-\frac{1}{2}}c^{-2}r^{-1}W'f(P'\pi\chi).$$

Substituting this in (29) and adding on the expression (27), we obtain the following value for the integral (26)

$$h^{-\frac{1}{2}}c^{-2}r^{-1}W'\{-\pi f(P'\pi\chi) + i\lambda(\pi\chi)\}e^{-iP'r/\hbar}. \quad (30)$$

Similarly the second term in the integrand in (25) gives

$$h^{-\frac{1}{2}}c^{-2}r^{-1}W'\{-\pi f(P'0\chi) - i\lambda(0\chi)\}e^{iP'r/\hbar}. \quad (31)$$

The sum of these two expressions is the value of  $(r\theta\phi|)$  when  $r$  is large.

We require that  $(r\theta\phi|)$  shall represent only outward moving particles, and hence it must be of the form of a multiple of  $e^{iP'r/\hbar}$ . Thus (30) must vanish, so that

$$\lambda(\pi\chi) = -i\pi f(P'\pi\chi). \quad (32)$$

We see in this way that the condition that  $(r\theta\phi|)$  shall represent only outward moving particles in the direction  $\theta = 0$  fixes the value of  $\lambda$  for the opposite direction  $\theta = \pi$ . Since the direction  $\theta = 0$  or  $\omega = 0$  of the pole of our polar coordinates is not in any way singular, we can generalize (32) to

$$\lambda(\omega\chi) = -i\pi f(P'\omega\chi), \quad (33)$$

which gives the value of  $\lambda$  for an arbitrary direction. This value substituted in (24) gives a result that may be written

$$(P\omega\chi|) = f(P\omega\chi)\{1/(W'-W) - i\pi\delta(W'-W)\}, \quad (34)$$



since one can substitute  $P'$  for  $P$  in the coefficient of a term involving  $\delta(W' - W)$  as a factor without changing the value of the term. *The condition that  $(P\omega\chi|)$  shall represent only outward moving particles is thus that it shall contain the factor*

$$\{1/(W' - W) - i\pi\delta(W' - W)\}. \quad (35)$$

It is interesting to note that this factor is of the form of the right-hand side of equation (17) of Chapter IV.

With  $\lambda$  given by (33), expression (30) vanishes and the value of  $(r0\phi|)$  for large  $r$  is given by expression (31) alone, thus

$$(r0\phi|) = -2\pi\hbar^{-1}c^{-2}r^{-1}W'f(P'0\chi)e^{iP'r/\hbar}.$$

This may be generalized to

$$(r\theta\phi|) = -2\pi\hbar^{-1}c^{-2}r^{-1}W'f(P'\omega\chi)e^{iP'r/\hbar},$$

giving the value of  $(r\theta\phi|)$  for any direction  $\theta$ ,  $\phi$  in terms of  $f(P'\omega\chi)$  for the same direction labelled by  $\omega$ ,  $\chi$ . This is of the form (12) with

$$u(\theta\phi) = -2\pi\hbar^{-1}c^{-2}W'f(P'\omega\chi)$$

and thus represents a distribution of outward moving particles of momentum  $P'$  whose number is

$$\frac{c^2 P'}{W'} |u|^2 = \frac{4\pi^2 W' P'}{\hbar c^2} |f(P'\omega\chi)|^2 \quad (36)$$

per unit solid angle per unit time. This distribution is the one represented by the  $(P\omega\chi|)$  of (34).

From this general result we can infer that, whenever we have a representative  $(P\omega\chi|)$  representing only outward moving particles and satisfying an equation of the type (23), the number per unit solid angle per unit time of these particles is given by (36). If this  $(P\omega\chi|)$  occurs in a problem in which the number of incident particles is one per unit volume, it will correspond to a scattering coefficient of amount

$$\frac{4\pi^2 W^0 W' P'}{\hbar c^4 P^0} |f(P'\omega\chi)|^2. \quad (37)$$

It is only the value of the function  $f(P\omega\chi)$  for the point  $P = P'$  that is of importance.

If we now apply this general theory to our equations (21) and (22), we have

$$f(P\omega\chi) = \hbar^{\frac{1}{2}}(P\omega\chi\alpha'|V|P^0\omega^0\chi^0\alpha^0).$$

Hence from (37) the scattering coefficient is

$$4\pi^2 \hbar^2 W^0 W' P' / c^4 P^0 \cdot |(P'\omega\chi\alpha'|V|P^0\omega^0\chi^0\alpha^0)|^2. \quad (38)$$



If one neglects relativity and puts  $W^0 W'/c^4 = m^2$ , this result reduces to the result (15) obtained in the preceding section by means of Green's theorem.

## 54. Dispersive Scattering

We shall now determine the scattering when the incident particle is capable of being absorbed, that is, when our unperturbed system of scatterer plus particle has closed stationary states with the particle absorbed. The existence of these closed states for the unperturbed system will be found to have a considerable effect on the scattering for the perturbed system, and indeed an effect that depends very much on the energy of the incident particle, giving rise to the phenomenon of dispersion in optics when the incident particle is taken to be a photon.

We use a representation for which the basic states are the stationary states of the unperturbed system, as was the case for the  $\mathbf{p}$ -representation of the preceding section. We take these stationary states to be the states  $\psi(\mathbf{p}'\alpha')$  for which the particle has a definite momentum  $\mathbf{p}'$  and the scatterer is in a definite state  $\alpha'$ , together with the closed states,  $\psi_k$  say, which form a separate discrete set, and assume that these states are all independent and orthogonal. This assumption is probably not justifiable when the particle is an electron or atomic nucleus, since in this case for an absorbed state  $\psi_k$  the particle will still certainly be somewhere, so that one would expect to be able to expand  $\psi_k$  in terms of the eigen- $\psi$ 's  $\psi(\mathbf{x}'\alpha')$  of  $x, y, z$ , and the  $\alpha$ 's, and hence also in terms of the  $\psi(\mathbf{p}'\alpha')$ . On the other hand, when the particle is a photon it will no longer exist for the absorbed states, which are then certainly independent of and orthogonal to the states  $\psi(\mathbf{p}'\alpha')$  for which the particle does exist. Thus the assumption is justified in this case, which is the important practical one.

The representative of a state will now consist of a discrete set of numbers ( $k|$ ) referring to the basic states  $\psi_k$  together with the three-dimensional continuous ranges of numbers ( $\mathbf{p}'\alpha'|$ ) referring to the  $\psi(\mathbf{p}'\alpha')$ , there being one such range for each set of values  $\alpha'$  for the  $\alpha$ 's. Similarly, the matrices representing dynamical variables will now contain discrete rows and columns labelled by  $k$  together with continuous ranges labelled by  $(\mathbf{p}, \alpha)$ . Thus, for example, the matrix representing  $V$ , the perturbing energy, will have elements  $(k'|V|k'')$ ,  $(k'|V|\mathbf{p}''\alpha'')$ ,  $(\mathbf{p}'\alpha'|V|k'')$ , and  $(\mathbf{p}'\alpha'|V|\mathbf{p}''\alpha'')$ .



Since we are concerned with scattering, we must still deal with *stationary* states of the whole system. We shall now, however, have to work to the second order of accuracy, so that we cannot use merely the first-order equation (3), but must use also (4). Equation (3) becomes, when written in terms of representatives in our present representation,

$$\left. \begin{aligned} \{W' - W\}(\mathbf{p}\alpha'|1) &= \sum_{\alpha''} \int (\mathbf{p}\alpha'|V|\mathbf{p}''\alpha'') \, d\mathbf{p}'' (\mathbf{p}''\alpha''|0) + \sum_{k''} (\mathbf{p}\alpha'|V|k'')(k''|0) \\ \{E - E_k\}(k|1) &= \sum_{\alpha''} \int (k|V|\mathbf{p}''\alpha'') \, d\mathbf{p}'' (\mathbf{p}''\alpha''|0) + \sum_{k''} (k|V|k'')(k''|0), \end{aligned} \right\} \quad (39)$$

where  $W'$  is the function of  $E$  and the  $\alpha'$ 's given by (18) and  $E_k$  is the energy of the stationary state  $\psi_k$  of the unperturbed system. Similarly, equation (4) becomes

$$\left. \begin{aligned} \{W' - W\}(\mathbf{p}\alpha'|2) &= \sum_{\alpha''} \int (\mathbf{p}\alpha'|V|\mathbf{p}''\alpha'') \, d\mathbf{p}'' (\mathbf{p}''\alpha''|1) + \sum_{k''} (\mathbf{p}\alpha'|V|k'')(k''|1) \\ \{E - E_k\}(k|2) &= \sum_{\alpha''} \int (k|V|\mathbf{p}''\alpha'') \, d\mathbf{p}'' (\mathbf{p}''\alpha''|1) + \sum_{k''} (k|V|k'')(k''|1). \end{aligned} \right\} \quad (40)$$

The unperturbed stationary state  $\psi_0$  will now be represented by

$$(\mathbf{p}\alpha|0) = h^{\frac{1}{2}} \delta_{\alpha\alpha^0} \delta(\mathbf{p} - \mathbf{p}^0) \quad (k|0) = 0, \quad (41)$$

instead of merely by (19), so (39) reduces to

$$\{W' - W\}(\mathbf{p}\alpha'|1) = h^{\frac{1}{2}} (\mathbf{p}\alpha'|V|\mathbf{p}^0\alpha^0) \quad (42)$$

$$\{E - E_k\}(k|1) = h^{\frac{1}{2}} (k|V|\mathbf{p}^0\alpha^0). \quad (43)$$

We may assume that the matrix elements  $(k'|V|k'')$  of  $V$  vanish, since these matrix elements are not essential to the phenomena under investigation, and if they did not vanish it would mean simply that the absorbed states  $\psi_k$  had not been suitably chosen. We shall further assume that the matrix elements  $(\mathbf{p}'\alpha'|V|\mathbf{p}''\alpha'')$  are of the second order of smallness when the matrix elements  $(k'|V|\mathbf{p}''\alpha'')$ ,  $(\mathbf{p}'\alpha'|V|k'')$  are taken to be of the first order of smallness. This assumption will be justified for the case of photons in Chapter XI. We now have from (43) and (42) that  $(k|1)$  is of the first order of smallness, provided  $E$  does not lie near one of the discrete set of energy-levels  $E_k$ , and  $(\mathbf{p}\alpha|1)$  is of the second order. The value of  $(\mathbf{p}\alpha|2)$  to the second order



will thus be given, from the first of equations (40), by

$$\{W' - W\}(\mathbf{p}_{\alpha'}|2) = h^{\frac{3}{2}} \sum_k (\mathbf{p}_{\alpha'}|V|k'')(k''|V|\mathbf{p}^0\alpha^0)/\{E - E_k\}.$$

The total correction to the second order, arising partly from  $(\mathbf{p}_{\alpha}|1)$  and partly from  $(\mathbf{p}_{\alpha}|2)$ , therefore satisfies

$$\begin{aligned} \{W' - W\}\{(\mathbf{p}_{\alpha'}|1) + (\mathbf{p}_{\alpha'}|2)\} \\ = h^{\frac{3}{2}}\{(\mathbf{p}_{\alpha'}|V|\mathbf{p}^0\alpha^0) + \sum_k (\mathbf{p}_{\alpha'}|V|k)(k|V|\mathbf{p}^0\alpha^0)/(E - E_k)\}. \end{aligned}$$

This equation is of the type (23), provided  $\alpha'$  is such that  $W' > mc^2$ , which means that  $\alpha'$  as a final state for the scatterer is not inconsistent with the law of conservation of energy. We can therefore infer from the general result (37) that the scattering coefficient is

$$\frac{4\pi^2 h^2 W^0 W' P'}{c^4 P^0} \left| (\mathbf{p}'_{\alpha'}|V|\mathbf{p}^0\alpha^0) + \sum_k \frac{(\mathbf{p}'_{\alpha'}|V|k)(k|V|\mathbf{p}^0\alpha^0)}{E - E_k} \right|^2. \quad (44)$$

The scattering may now be considered as composed of two parts, a part that arises from the matrix element  $(\mathbf{p}'_{\alpha'}|V|\mathbf{p}^0\alpha^0)$  of the perturbing energy and a part that arises from the matrix elements  $(\mathbf{p}'_{\alpha'}|V|k)$  and  $(k|V|\mathbf{p}^0\alpha^0)$ . The first part, which is the same as our previously obtained result (38), may be called the true scattering. The second part may be considered as arising from an absorption of the incident particle into some state  $k$ , followed immediately by a re-emission in a different direction. The fact that we have to add the two terms before taking the square of the modulus denotes interference between the two kinds of scattering. There is no experimental way of separating the two kinds, the distinction between them being only mathematical.

## 55. Resonance Scattering

Suppose the energy of the incident particle to be varied continuously while the initial state  $\alpha^0$  of the scatterer is kept fixed, so that the total energy  $E$  varies continuously. The formula (44) now shows that as  $E$  approaches one of the discrete set of energy-levels  $E_k$ , the scattering becomes very large. In fact, according to formula (44) the scattering should be infinite when  $E$  is exactly equal to an  $E_k$ . An infinite scattering coefficient is, of course, physically impossible, so that we can infer that the approximations used in deriving (44) are no longer legitimate when  $E$  is close to an  $E_k$ . To investigate the scattering in this case we must therefore go back to the exact equation

$$\{E - H_0\}\psi = V\psi,$$



which is the same as (2) of Chapter VIII, and use a different method of approximating to its solution. This exact equation, written in terms of representatives, becomes

$$\left. \begin{aligned} \{W' - W\}(\mathbf{p}\alpha') &= \sum_{\alpha''} \int (\mathbf{p}\alpha' | V | \mathbf{p}''\alpha'') d\mathbf{p}'' (\mathbf{p}''\alpha'' |) + \sum_{k''} (\mathbf{p}\alpha' | V | k'')(k'' |) \\ \{E - E_k\}(k |) &= \sum_{\alpha''} \int (k | V | \mathbf{p}''\alpha'') d\mathbf{p}'' (\mathbf{p}''\alpha'' |) + \sum_{k''} (k | V | k'')(k'' |). \end{aligned} \right\} \quad (45)$$

Let us take one particular  $E_k$  and consider the case when  $E$  is close to it. The large term in the scattering coefficient (44) now arises from those elements of the matrix representing  $V$  that lie in row  $k$  or in column  $k$ , i.e. those of the type  $(k | V | \mathbf{p}\alpha)$  or  $(\mathbf{p}\alpha | V | k)$ . The scattering arising from the other matrix elements of  $V$  is of a smaller order of magnitude. This suggests that in our exact equations (45) we should make the approximation of neglecting all the matrix elements of  $V$  except the important ones, which are those of the type  $(\mathbf{p}\alpha' | V | k)$  or  $(k | V | \mathbf{p}\alpha')$ , where  $\alpha'$  is a state of the scatterer that has not too much energy to be disallowed as a final state by the law of conservation of energy. These equations then reduce to

$$\left. \begin{aligned} \{W' - W\}(\mathbf{p}\alpha') &= (\mathbf{p}\alpha' | V | k)(k |) \\ \{E - E_k\}(k |) &= \sum_{\alpha'} \int (k | V | \mathbf{p}\alpha') d\mathbf{p} (\mathbf{p}\alpha' |), \end{aligned} \right\} \quad (46)$$

the  $\alpha'$  summation being over those values of  $\alpha'$  for which  $W'$  given by (18) is  $> mc^2$ . These equations are now sufficiently simple for us to be able to solve exactly without further approximation.

From the first of equations (46) we obtain by division

$$(\mathbf{p}\alpha' |) = (\mathbf{p}\alpha' | V | k)(k |) / \{W' - W\} + \lambda \delta(W' - W). \quad (47)$$

We must choose  $\lambda$ , which may be any function of the momentum  $\mathbf{p}$  and  $\alpha'$ , such that (47) represents the incident particles (19) together with only outward moving particles. [The right-hand side of (19), with  $\alpha'$  substituted for  $\alpha$ , is actually of the form  $\lambda \delta(W' - W)$ , since the conditions  $\alpha' = \alpha^0$  and  $\mathbf{p} = \mathbf{p}^0$  for this right-hand side not to vanish lead to  $W' = E - H_s(\alpha') = E - H_s(\alpha^0) = W^0$  and  $W = W^0$ , which together give  $W' = W$ .] Thus (47) must be

$$(\mathbf{p}\alpha' |) = h^{\frac{3}{2}} \delta_{\alpha'\alpha^0} \delta(\mathbf{p} - \mathbf{p}^0) + (\mathbf{p}\alpha' | V | k)(k |) \{1/(W' - W) - i\pi \delta(W' - W)\}, \quad \dots \quad (48)$$

and from the general formula (37) the scattering coefficient will be

$$4\pi^2 W^0 W' P' / hc^4 P^0 \cdot |(\mathbf{p}'\alpha' | V | k)|^2 |(k |)|^2. \quad (49)$$



It remains for us to determine the value of  $(k|)$ . We can do this by substituting for  $(\mathbf{p}'_\alpha|)$  in the second of equations (46) its value given by (48). This gives

$$\begin{aligned} & \{E - E_k\}(k|) \\ &= h^{\frac{1}{2}}(k|V|\mathbf{p}^0_{\alpha^0}) + (k|) \sum_{\alpha'} \int |(k|V|\mathbf{p}'_{\alpha'})|^2 \{1/(W' - W) - i\pi\delta(W' - W)\} d\mathbf{p} \\ &= h^{\frac{1}{2}}(k|V|\mathbf{p}^0_{\alpha^0}) + (k|)\{a - ib\}, \end{aligned}$$

where 
$$a = \sum_{\alpha'} \int |(k|V|\mathbf{p}'_{\alpha'})|^2 d\mathbf{p} / (W' - W) \quad (50)$$

$$\begin{aligned} \text{and } b &= \pi \sum_{\alpha'} \int |(k|V|\mathbf{p}'_{\alpha'})|^2 \delta(W' - W) d\mathbf{p} \\ &= \pi \sum_{\alpha'} \iiint |(k|V|P\omega\chi\alpha')|^2 \delta(W' - W) P^2 dP \sin\omega d\omega d\chi \\ &= \pi \sum_{\alpha'} P' W' c^{-2} \iint |(k|V|P'\omega\chi\alpha')|^2 \sin\omega d\omega d\chi. \end{aligned} \quad (51)$$

$$\text{Thus} \quad (k|) = h^{\frac{1}{2}}(k|V|\mathbf{p}^0_{\alpha^0}) / \{E - E_k - a + ib\}. \quad (52)$$

Note that  $a$  and  $b$  are real and that  $b$  is positive.

This value for  $(k|)$  substituted in (49) gives for the scattering coefficient

$$\frac{4\pi^2 h^2 W^0 W' P'}{c^4 P^0} \frac{|(\mathbf{p}'_{\alpha'}|V|k)|^2 |(k|V|\mathbf{p}^0_{\alpha^0})|^2}{(E - E_k - a)^2 + b^2}. \quad (53)$$

One can obtain the total effective area that the incident particle must hit in order to be scattered anywhere by integrating (53) over all directions of scattering, i.e. by integrating over all directions of the vector  $\mathbf{p}'$  with its magnitude kept fixed at  $P'$ , and then summing over all  $\alpha'$  that are to be taken into consideration, i.e. for which  $W' > mc^2$ . This gives, with the help of (51), the result

$$\frac{4\pi h^2 W^0}{c^2 P^0} \frac{b |(k|V|\mathbf{p}^0_{\alpha^0})|^2}{(E - E_k - a)^2 + b^2}. \quad (54)$$

If we suppose  $E$  to vary continuously through the value  $E_k$ , the main variation of (53) or (54) will be due to the small denominator  $(E - E_k - a)^2 + b^2$ . If we neglect the dependence of the other factors in (53) and (54) on  $E$ , then the maximum scattering will occur when  $E$  has the value  $E_k + a$  and the scattering will be half its maximum when  $E$  differs from this value by an amount  $b$ . The large amount of scattering that occurs for values of the energy of the incident particle that make  $E$  nearly equal to  $E_k$  give rise to the phenomenon of an absorption line. The centre of the line is displaced by an amount



$a$  from the resonance energy of the incident particle, i.e. the energy which would make the total energy just  $E_k$ , while the quantity  $b$  is what is sometimes called the half-width of the line.

## 56. Emission and Absorption

For studying emission and absorption we must consider non-stationary states of the system and must use the perturbation method of § 47. To determine the coefficient of spontaneous emission we must take an initial state for which the particle is absorbed, so that the representative of the state is

$$(k|) = 1 \quad (p\alpha|) = 0,$$

and determine the probability that at some later time the particle shall be on its way to infinity with a definite momentum. The method of § 49 can now be applied. From the result (31) of that section we see that the probability per unit time per unit range of  $\omega$  and  $\chi$  of the particle being emitted in any direction  $\omega'$ ,  $\chi'$  with the scatterer being left in state  $\alpha'$  is

$$2\pi\hbar^{-1}|(k|V|W'\omega'\chi'\alpha')|^2, \quad (55)$$

provided, of course, that  $\alpha'$  is such that the energy  $W'$ , given by (18), of the particle is greater than  $mc^2$ . For values of  $\alpha'$  that do not satisfy this condition there is no emission possible. The matrix element  $(k|V|W'\omega'\chi'\alpha')$  here must refer to a representation in which  $W$ ,  $\omega$ ,  $\chi$  and  $\alpha$  are diagonal with the weight function unity. The matrix elements of  $V$  appearing in the three preceding sections refer to a representation in which  $p_x, p_y, p_z$  are diagonal with the weight function unity, or  $P$ ,  $\omega$ ,  $\chi$  are diagonal with the weight function  $P^2\sin\omega$ . They would thus refer to a representation in which  $W, \omega, \chi$  are diagonal with the weight function  $dP/dW \cdot P^2\sin\omega = WP/c^2 \cdot \sin\omega$ . Thus the matrix element  $(k|V|W'\omega'\chi'\alpha')$  in (55) is equal to  $(W'P'/c^2 \cdot \sin\omega')^{\frac{1}{2}}$  times our previous matrix element  $(k|V|P'\omega'\chi'\alpha')$  or  $(k|V|\mathbf{p}'\alpha')$ , so that (55) is equal to

$$\frac{2\pi}{\hbar} \frac{W'P'}{c^2} \sin\omega' |(k|V|\mathbf{p}'\alpha')|^2.$$

The probability of emission per unit solid angle per unit time, with the scatterer simultaneously dropping to state  $\alpha'$ , is thus

$$\frac{2\pi}{\hbar} \frac{W'P'}{c^2} |(k|V|\mathbf{p}'\alpha')|^2. \quad (56)$$

To obtain the total probability per unit time of the particle being emitted in any direction, with any final state for the scatterer, we



must integrate (56) over all angles  $\omega'$ ,  $\chi'$  and sum over all states  $\alpha'$  whose energy  $H_s(\alpha')$  is such that  $H_s(\alpha') + mc^2 < E_k$ . The result is just  $2b/\hbar$ , where  $b$  is defined by (51). *There is thus this simple relation between the total emission coefficient and the half-breadth  $b$  of the absorption line.*

Let us now consider absorption. This requires that we shall take an initial state for which the particle is certainly not absorbed but is incident with a definite momentum. Thus the representative of the initial state must be of the form (41). We must now determine the probability of the particle being absorbed after time  $T$ . Since our final state  $\psi_k$  is not one of a continuous range, we cannot use directly the result (31) of § 49. If, however, we take

$$(\mathbf{p}\alpha|)_0 = \delta_{\alpha\alpha^0} \delta(\mathbf{p}-\mathbf{p}^0) \quad (k|)_0 = 0 \quad (57)$$

as the representative of the initial state, the analysis of §§ 47 and 49 is still applicable as far as equation (28) and shows us that the probability of the particle being absorbed into state  $\psi_k$  after time  $T$  is

$$2|(k|V|\mathbf{p}^0\alpha^0)|^2[1 - \cos\{(E_k - E)T/\hbar\}]/(E_k - E)^2.$$

This corresponds to a distribution of incident particles of density  $h^{-3}$ , owing to the omission of the factor  $h^3$  from (57), as compared with (41). The probability of there being an absorption after time  $T$  when there is one incident particle crossing unit area per unit time is therefore

$$2h^3 W^0/c^2 P^0 \cdot |(k|V|\mathbf{p}^0\alpha^0)|^2[1 - \cos\{(E_k - E)T/\hbar\}]/(E_k - E)^2. \quad (58)$$

To obtain the absorption coefficient we must consider the incident particles not all to have exactly the same energy  $W^0 = E - H_s(\alpha^0)$ , but to have a distribution of energy values about the correct value  $E_k - H_s(\alpha^0)$  required for absorption. If we take a beam of incident particles consisting of one crossing unit area per unit time per unit energy range, the probability of there being an absorption after time  $T$  will be given by the integral of (58) with respect to  $E$ . This integral may be evaluated in the same way as (29) of § 49 and is equal to

$$4\pi^2 h^2 W^0 T/c^2 P^0 \cdot |(k|V|\mathbf{p}^0\alpha^0)|^2.$$

The probability per unit time of an absorption taking place with an incident beam of one particle per unit area per unit time per unit energy range is therefore

$$4\pi^2 h^2 W^0/c^2 P^0 \cdot |(k|V|\mathbf{p}^0\alpha^0)|^2, \quad (59)$$

which is the absorption coefficient.



The connexion between the absorption and emission coefficients (59) and (56) and the resonance scattering coefficients calculated in the preceding section should be noted. When the incident beam does not consist of particles all with the same energy, but consists of a unit distribution of particles per unit energy range crossing unit area per unit time, the total number of incident particles with energies near an absorption line that get scattered will be given by the integral of (54) with respect to  $E$ . If one neglects the dependence of the numerator of (54) on  $E$ , this integral will, since

$$\int_{-\infty}^{\infty} \frac{b}{(E - E_k - a)^2 + b^2} dE = \pi,$$

have just the value (59). Thus *the total number of scattered particles in the neighbourhood of an absorption line is equal to the total number absorbed*. We can therefore regard all these scattered particles as absorbed particles that are subsequently re-emitted in a different direction. Further, the number of particles in the neighbourhood of the absorption line that get scattered per unit solid angle about a given direction specified by  $\mathbf{p}'$  and then belong to scatterers in state  $\alpha'$  will be given by the integral with respect to  $E$  of (53), which integral has in the same way the value

$$\frac{4\pi^2 \hbar^2 W^0 W' P'}{c^4 P^0} \frac{\pi}{b} |(\mathbf{p}' \alpha' | V | k)|^2 |(k | V | \mathbf{p}^0 \alpha^0)|^2.$$

This is just equal to the absorption coefficient (59) multiplied by the emission coefficient (56) divided by  $2b/\hbar$ , the total emission coefficient. This is in agreement with the point of view of regarding the resonance scattered particles as those that are absorbed and then re-emitted, with the absorption and emission processes governed independently each by its own probability law, since this point of view would make the fraction of the total number of absorbed particles that are re-emitted in a unit solid angle about a given direction just the emission coefficient for this direction divided by the total emission coefficient.



## SYSTEMS CONTAINING SEVERAL SIMILAR PARTICLES

## 57. Symmetrical and Antisymmetrical States

IF a system in atomic physics contains a number of particles of the same kind, e.g. a number of electrons, the particles are absolutely indistinguishable one from another. No observable change is made when two of them are interchanged. This circumstance gives rise to some curious phenomena in quantum mechanics having no analogue in the classical theory, which arise from the fact that in quantum mechanics a transition may occur resulting in merely the interchange of two similar particles, which transition then could not be detected by any observational means. A satisfactory theory ought, of course, to count two observationally indistinguishable states as the same state and to deny that any transition does occur when two similar particles exchange places. We shall find that such a theory can be developed in agreement with the principles of quantum mechanics.

Suppose we have a system containing  $n$  similar particles. We may take as our dynamical variables a set of variables  $\xi_1$  describing the first particle, the corresponding set  $\xi_2$  describing the second particle, and so on up to the set  $\xi_n$  describing the  $n$ th particle. We shall then have the  $\xi_r$ 's commuting with the  $\xi_s$ 's for  $r \neq s$ . (We may require certain extra variables, describing what the system consists of in addition to the  $n$  similar particles, but it is not necessary to mention these explicitly in the present chapter.) The Hamiltonian describing the motion of the system will now be expressible as a function of the  $\xi_1, \xi_2, \dots, \xi_n$ . The fact that the particles are similar requires that *the Hamiltonian shall be a symmetrical function of the  $\xi_1, \xi_2, \dots, \xi_n$ , i.e. it shall remain unchanged when the sets of variables  $\xi_r$  are interchanged or permuted in any way.* This condition must hold, no matter what perturbations are applied to the system.

We may take a representation with observables  $q_1, q_2, \dots, q_n$  diagonal, which are such that the  $q_1$ 's are a complete set of commuting observables describing the first particle, the  $q_2$ 's are the corresponding observables describing the second particle, and so on. We may further choose the phases of the representation in the same way for each of the particles. (This means, for example, that if a certain momentum  $p_1$  describing the first particle is represented by  $-i\hbar\partial/\partial q_1$ ,



the corresponding momentum  $p_r$  describing the  $r$ -th particle must be represented by  $-i\hbar\partial/\partial q_r$ .) The representation will then treat all the particles on the same footing. The condition that the Hamiltonian  $H$  is symmetrical between all the particles may now be expressed by the condition that its representative  $(q'_1 q'_2 \dots q'_n | H | q''_1 q''_2 \dots q''_n)$ , or  $(q' | H | q'')$  for brevity, is symmetrical between all the  $q$ 's, i.e. that it remains unchanged if any permutation is applied to the  $q'$ 's and the same permutation to the  $q''$ 's. This condition may be expressed analytically thus,

$$(q' | H | q'') = (Pq' | H | Pq''), \quad (1)$$

where  $P$  denotes any permutation of the numbers  $1, 2, \dots, n$  and  $Pq'$  denotes the set of numbers obtained by applying the permutation  $P$  to the suffixes of  $q'_1, q'_2, \dots, q'_n$ .

Let  $(q'_1 q'_2 \dots q'_n |)$  or  $(q' |)$  be the wave function representing any state. It will satisfy the wave equation

$$i\hbar \frac{d}{dt} (q' |) = \int (q' | H | q'') dq'' (q'' |). \quad (2)$$

If we apply any permutation  $P$  to the variables  $q'$  in  $(q' |)$  we shall obtain a function  $(Pq' |)$  satisfying

$$\begin{aligned} i\hbar \frac{d}{dt} (Pq' |) &= \int (Pq' | H | q'') dq'' (q'' |) \\ &= \int (Pq' | H | Pq'') dq'' (Pq'' |), \end{aligned}$$

since we can apply any permutation to the variables of integration  $q''$  in the integrand without changing the value of the integral. With the help of (1) this becomes

$$i\hbar \frac{d}{dt} (Pq' |) = \int (q' | H | q'') dq'' (Pq'' |), \quad (3)$$

which shows that  $(Pq' |)$  is a solution of the wave equation (2). Hence *if we apply any permutation to the variables in a solution of the wave equation we obtain another solution.*

Suppose we take a wave function  $(q' |)$  which, at some particular time  $t$ , is a symmetrical function of all the  $q'$ 's, so that

$$(q' |) = (Pq' |) \quad (4)$$

for any  $P$ . The right-hand sides of (2) and (3) are now equal, so that

$$\frac{d}{dt} (q' |) = \frac{d}{dt} (Pq' |).$$

This equation is the time derivative of (4) and shows that if (4) holds



at one particular time it holds also at a slightly later time, and thus by induction it holds at all times. Thus *if a wave function is initially symmetrical it always remains symmetrical*.

Similarly, we may take a wave function  $(q'|)$  which, at some particular time, is antisymmetrical, i.e.  $(q'_1 q'_2 \dots q'_n|)$  changes sign with interchange of any pair of  $q'$ 's. We shall then have

$$(q'|) = \pm(Pq'|), \quad (5)$$

the  $+$  or  $-$  sign being taken according to whether the permutation  $P$  is even or odd (i.e. according to whether  $P$  can be built up from an even or an odd number of simple interchanges). The same argument as before now shows that *if a wave function is initially antisymmetrical it always remains antisymmetrical*.

Let us make a canonical transformation to a  $Q$ -representation which, like the original  $q$ -representation, treats all the particles on the same footing. This means that the  $Q$ 's consist of corresponding sets of observables  $Q_1, Q_2, \dots, Q_n$  describing the first, second, ...,  $n$ -th particle respectively and that the phases are chosen in the same way for each of the particles. The transformation function will now, from (11) of Chapter V, be of the form

$$(Q'_1 Q'_2 \dots Q'_n | q'_1 q'_2 \dots q'_n) = (Q'_1 | q'_1)(Q'_2 | q'_2) \dots (Q'_n | q'_n), \quad (6)$$

in which each factor  $(Q'_r | q'_r)$  is the same function of its variables  $Q'_r, q'_r$ . This condition gives, if we denote  $(Q'_1 Q'_2 \dots Q'_n | q'_1 q'_2 \dots q'_n)$  by  $(Q' | q')$  for brevity,

$$(Q' | q') = (PQ' | Pq'), \quad (7)$$

for an arbitrary permutation  $P$ . The new wave function representing any state is given by

$$(Q' |) = \int (Q' | q') dq' (q' |). \quad (8)$$

From this equation we can deduce that

$$\begin{aligned} (PQ' |) &= \int (PQ' | q') dq' (q' |) \\ &= \int (PQ' | Pq') dq' (Pq' |) \\ &= \int (Q' | q') dq' (Pq' |) \end{aligned} \quad (9)$$

with the help of (7). Now if  $(q' |)$  is symmetrical, so that equation (4) holds, the right-hand sides of (8) and (9) are equal. We then have  $(Q' |) = (PQ' |)$ , so that  $(Q' |)$  is also symmetrical. Similarly, if  $(q' |)$  is antisymmetrical,  $(Q' |)$  is also antisymmetrical. Thus *the property of the representative of a state of being symmetrical or antisymmetrical*



remains invariant under a transformation of the coordinate system. This invariance shows that the property of being symmetrical or anti-symmetrical is a property of the states themselves and not merely a property of their representatives. Thus we can talk about symmetrical and antisymmetrical states. Our preceding result shows that if a state is initially symmetrical or antisymmetrical, it always remains so.

The invariance and permanence of the symmetry properties of the states means that for some particular kind of particle it is quite possible for only symmetrical or only antisymmetrical states to occur in nature. Whether this is the case cannot be decided by any general theoretical considerations, but can be settled only by reference to special experimentally determined facts about the particles in question. For photons one can settle the question by making use of Planck's radiation law. Only when one assumes *the symmetrical states for photons* does one get a statistical mechanics leading to Planck's law for radiation in statistical equilibrium. This statistical mechanics is known as the *Einstein-Bose statistics*, as it was first introduced by Bose and Einstein before the arrival of the modern quantum mechanics.

For electrons we use the fact that, if we make the approximation of regarding the electrons in an atom as each moving in its own 'orbit' (i.e. as being each describable by its own wave function involving only its own variables), then no two electrons will ever be in the same orbit. This fact, which is known as *Pauli's exclusion principle*, may be inferred from general experimental evidence on atomic structure. Let us see how to fit it in with the theory. If the wave functions representing the different electronic orbits are

$$(q'|\alpha_1), (q'|\alpha_2), \dots, (q'|\alpha_n),$$

a wave function representing the whole atom will be given by the product

$$(q'|\alpha_1)(q'|\alpha_2)\dots(q'|\alpha_n) = (q'|\alpha) \quad (10)$$

say, for brevity. Other wave functions representing the same distribution of electrons over the various orbits may be obtained by applying any permutation to the  $\alpha$ 's in (10). There will be altogether  $n!$  such wave functions, the general one being  $(q'|P\alpha)$ . Any linear combination of these wave functions will also represent the same electron distribution. One such linear combination is the sum

$$\sum_P (q'|P\alpha), \quad (11)$$



which is symmetrical between all the  $q$ 's. Another is

$$\sum_P \pm (q' | P \alpha), \quad (12)$$

the  $+$  or  $-$  sign being taken according to whether  $P$  is an even or odd permutation, and this one is antisymmetrical. The antisymmetrical wave function (12) has the property that it vanishes identically if two of the  $\alpha$ 's are equal. Hence if we assume that *for electrons only antisymmetrical states occur*, we shall get the result that there are no states with two electrons in the same orbit, which is just Pauli's exclusion principle. This assumption is the only one we can make which will lead to Pauli's exclusion principle.

In this way we can see that for photons we must take the symmetrical states and for electrons the antisymmetrical states. When only the symmetrical or only the antisymmetrical states are allowed for a particular kind of particle, the theory can no longer make a distinction between two states which differ only through a permutation of the particles, so that the difficulties mentioned at the beginning of this section disappear.

## 58. Permutations as Dynamical Variables

Let us now build up a general theory for a system containing  $n$  similar particles when states with any kind of symmetry properties are allowed, i.e. when there is no restriction to only symmetrical or only antisymmetrical states. The general state now will not be symmetrical or antisymmetrical, nor will it be expressible linearly in terms of symmetrical and antisymmetrical states when  $n > 2$ .

If  $P$  denotes any permutation and  $\psi$  any  $\psi$ -vector, we can give a meaning to  $P\psi$ , the  $\psi$ -vector obtained by operating on  $\psi$  with  $P$ . We define  $P\psi$  to be the  $\psi$ -vector whose representative is  $(Pq' |)$ , obtained by applying the permutation  $P$  to the representative  $(q' |)$  of  $\psi$ . This  $P\psi$  is independent of the representation used for defining it, as follows from equation (9). Further, the operation by which  $P\psi$  is obtained from  $\psi$  is a linear one. Hence we can regard  $P\psi$  as the product of a dynamical variable  $P$  with  $\psi$ , i.e. *we can regard the permutation  $P$  as a dynamical variable*.

There are  $n!$  permutations, each of which can be regarded as a dynamical variable. One of them,  $P_1$  say, is the identical permutation, which is equal to unity. If  $\psi$  denotes a symmetrical state, we have

$$P\psi = \psi \quad (13)$$



for any  $P$ , and hence a symmetrical  $\psi$  is an eigen- $\psi$  of every permutation belonging to the eigenvalue unity. Similarly, an antisymmetrical  $\psi$  is an eigen- $\psi$  of every permutation belonging to the eigenvalue  $\pm 1$  according to whether the permutation is even or odd. The product of any two permutations is a third permutation and hence any function of the permutations is reducible to a linear function of them. Any permutation  $P$  has a reciprocal  $P^{-1}$  satisfying

$$PP^{-1} = P^{-1}P = P_1 = 1.$$

A permutation  $P$ , like any other dynamical variable, can be represented by a matrix. Its  $q$ -representative  $(q'|P|q'')$  will satisfy

$$\int (q'|P|q'') dq'' (q''|) = (Pq'|)$$

and hence

$$(q'|P|q'') = \delta(Pq' - q'') \quad (14)$$

$$= \delta(q' - P^{-1}q''). \quad (15)$$

The  $\delta$  function in (14) or (15) denotes the product of  $n$  factors of the type  $\delta(\{Pq'_r\} - q''_r)$  or  $\delta(q'_r - \{P^{-1}q''_r\})$  respectively. The conjugate complex of  $P$  is given by

$$\begin{aligned} (q'|\bar{P}|q'') &= \overline{(q''|P|q')} = \delta(q'' - P^{-1}q') \\ &= (q'|P^{-1}|q'') \end{aligned}$$

from (15) and (14), so that

$$\bar{P} = P^{-1}. \quad (16)$$

Thus a permutation is not in general a real dynamical variable, its conjugate complex being equal to its reciprocal.

Any permutation of the numbers 1, 2, 3, ...,  $n$  may be expressed in the cyclic notation, e.g. with  $n = 8$

$$P_a = (143)(27)(58)(6), \quad (17)$$

in which each number is to be replaced by the succeeding number in a bracket, unless it is the last in a bracket, when it is to be replaced by the first in that bracket. Thus  $P_a$  changes the numbers 12345678 into 47138625. The type of any permutation is specified by the partition of the number  $n$  which is provided by the number of numbers in each of the brackets. Thus the type of  $P_a$  is specified by the partition  $8 = 3 + 2 + 2 + 1$ . Permutations of the same type, i.e. corresponding to the same partition, we shall call *similar*. Thus, for example,  $P_a$  in (17) is similar to

$$P_b = (871)(35)(46)(2). \quad (18)$$



The whole of the  $n!$  possible permutations may be divided into sets of similar permutations, each such set being called a *class*. The permutation  $P_1 = 1$  forms a class by itself. Any permutation is similar to its reciprocal.

When two permutations  $P_a$  and  $P_b$  are similar, either of them  $P_b$  may be obtained by making a certain permutation  $P$  in the other  $P_a$ . Thus, in our example (17), (18) we can take  $P$  to be the permutation that changes 14327586 into 87135462, i.e. the permutation

$$P = (18623)(475).$$

We then have the algebraic relation between  $P_a$  and  $P_b$

$$P_b = PP_a P^{-1}. \quad (19)$$

To verify this, we observe that the product  $P_a \psi$  of  $P_a$  with any  $\psi$  is changed into  $P_b \psi$  if one applies the permutation  $P$  to the  $P_a$  in the product but not to the  $\psi$ . If we multiply the product by  $P$  on the left, we are applying this permutation to the whole  $\psi$ -symbol  $P_a \psi$  and thus to both the  $P_a$  and the  $\psi$ , so that we must insert another factor  $P^{-1}$  between the  $P_a$  and the  $\psi$ , giving us  $PP_a P^{-1} \psi$  to equate to  $P_b \psi$ . An alternative proof consists in noting that when the permutation  $P$  is applied to the representative  $\delta(P_a q' - q'')$  of  $P_a$ , it gives  $\delta(PP_a q' - P q'')$  or  $\delta(PP_a P^{-1} q' - q'')$ , which is just the representative of  $PP_a P^{-1}$ .

Equation (19) is the general formula showing when two permutations  $P_a$  and  $P_b$  are similar. Of course  $P$  is not uniquely determined when  $P_a$  and  $P_b$  are given, but the existence of any  $P$  satisfying (19) is sufficient to show that  $P_a$  and  $P_b$  are similar.

## 59. Permutations as Constants of the Motion

Let us see how one of our permutation dynamical variables  $P$  varies with the time. The fact that the Hamiltonian is symmetrical leads at once to the equation

$$PH = HP, \quad (20)$$

as may be verified by a similar argument to that used for equation (19), or alternatively by a direct application of the matrix representatives. Thus from (14)

$$(q' | PH | q'') = \int \delta(Pq' - q''') dq''' (q''' | H | q'') = (Pq' | H | q'')$$

and from (15)

$$(q' | HP | q'') = \int (q' | H | q''') dq''' \delta(q''' - P^{-1}q'') = (q' | H | P^{-1}q''),$$



and the two right-hand sides are now equal from (1). Equation (20) shows that *each permutation is a constant of the motion*. The  $P$ 's are still constants when arbitrary perturbations are applied to the system, provided the perturbing energy to be added to the Hamiltonian is symmetrical. Thus the constancy of the  $P$ 's is perfect.

In dealing with any system in quantum mechanics, when we have found a constant of the motion  $\alpha$ , we know that if for any state of motion,  $\alpha$  initially has the numerical value  $\alpha'$ , then it always has this value, so that we can assign different numbers  $\alpha'$  to the different states and so obtain a useful classification of the states. The procedure is not so straightforward, however, when we have several constants of the motion  $\alpha$  which do not commute (as is the case with our permutations  $P$ ), since we cannot assign numerical values for all the  $\alpha$ 's simultaneously to any state. Let us first take the case of a system whose Hamiltonian does not involve the time explicitly. The existence of constants of the motion  $\alpha$  which do not commute is then a sign that the system is degenerate. This is because, for a non-degenerate system, the Hamiltonian  $H$  by itself forms a complete set of commuting observables and hence, from the theorem at the top of page 60, each of the  $\alpha$ 's is a function of  $H$  and therefore commutes with any other  $\alpha$ .

We must now look for a function  $\beta$  of the  $\alpha$ 's which has one and the same numerical value  $\beta'$  for all those states belonging to one energy-level  $H'$ , so that we can use  $\beta$  for classifying the energy-levels of the system. We can express the condition for  $\beta$  by saying that it must be a function of  $H$  and must therefore commute with every dynamical variable that commutes with  $H$ , i.e. with every constant of the motion. If the  $\alpha$ 's are the only constants of the motion, or if they are a set that commute with all other independent constants of the motion, our problem reduces to finding a function  $\beta$  of the  $\alpha$ 's which commutes with all the  $\alpha$ 's. We can then assign a numerical value  $\beta'$  for  $\beta$  to each energy-level of the system. If we can find several such functions  $\beta$ , they must all commute with each other, so that we can give them all numerical values simultaneously and obtain a complete classification of the energy-levels. When the Hamiltonian involves the time explicitly one cannot talk about energy-levels, but the  $\beta$ 's will still give a useful classification for the states.

We follow this method in dealing with our permutations  $P$ . We must find a function  $\chi$  of the  $P$ 's such that  $P\chi P^{-1} = \chi$  for every  $P$ . It is evident that a possible  $\chi$  is  $\sum P_c$ , the sum of all the permutations



in a certain class  $c$ , i.e. the sum of a set of similar permutations, since  $\sum PP_c P^{-1}$  must consist of the same permutations summed in a different order. There will be one such  $\chi$  for each class. Further, there can be no other independent  $\chi$ , since an arbitrary function of the  $P$ 's can be expressed as a linear function of them with numerical coefficients, and it will not then commute with every  $P$  unless the coefficients of similar  $P$ 's are always the same. We thus obtain all the  $\chi$ 's that can be used for classifying the states. It is convenient to define each  $\chi$  as an average instead of a sum, thus

$$\chi_c = n_c^{-1} \sum P_c,$$

where  $n_c$  is the number of  $P$ 's in the class  $c$ . An alternative expression for  $\chi_c$  is

$$\chi_c = n!^{-1} \sum_P PP_c P^{-1}, \quad (21)$$

the sum being extended over all the  $n!$  permutations  $P$ , it being easy to verify that this sum contains each member of the class  $c$  the same number of times. For each permutation  $P$  there is one  $\chi$ ,  $\chi(P)$  say, equal to the average of all permutations similar to  $P$ . One of the  $\chi$ 's is  $\chi(P_1) = 1$ .

The constants of the motion  $\chi_1, \chi_2, \dots, \chi_m$  obtained in this way will each have a definite numerical value for every stationary state of the system, in the case when the Hamiltonian does not involve the time explicitly, and also in the general case can be used for classifying the states, there being one set of states for every permissible set of numerical values  $\chi'_1, \chi'_2, \dots, \chi'_m$  for the  $\chi$ 's. Since the  $\chi$ 's are perfect constants of the motion, these sets of states will be *exclusive*, i.e. transitions will never take place from a state in one set to a state in another.

The permissible sets of values  $\chi'$  that one can give to the  $\chi$ 's are limited by the fact that there exist algebraic relations between the  $\chi$ 's. The product of any two  $\chi$ 's,  $\chi_p \chi_q$ , is of course expressible as a linear function of the  $P$ 's, and since it commutes with every  $P$  it must be expressible as a linear function of the  $\chi$ 's, thus

$$\chi_p \chi_q = a_1 \chi_1 + a_2 \chi_2 + \dots + a_m \chi_m, \quad (22)$$

where the  $a$ 's are numbers. Any numerical values  $\chi'$  that one gives to the  $\chi$ 's must be eigenvalues of the  $\chi$ 's and must satisfy these same algebraic equations. For every solution  $\chi'$  of these equations there is one exclusive set of states. One solution is evidently  $\chi'_p = 1$  for every  $\chi_p$ , giving the set of symmetrical states satisfying (13). A second



obvious solution, giving the set of antisymmetrical states, is  $\chi'_p = \pm 1$ , the  $+$  or  $-$  sign being taken according to whether the permutations in the class  $p$  are even or odd. The other solutions may be worked out in any special case by ordinary algebraic methods, as the coefficients  $a$  in (22) may be obtained directly by a consideration of the types of permutation to which the  $\chi$ 's concerned refer. Any solution is, apart from a certain factor, what is called in group theory a *character* of the group of permutations. The  $\chi$ 's are all real dynamical variables, since each  $P$  and its conjugate complex  $P^{-1}$  are similar and will occur added together in the definition of any  $\chi$ , so that the  $\chi$ 's must be all real numbers.

The number of possible solutions of the equations (22) may easily be determined, since it must equal the number of different eigenvalues of an arbitrary function  $B$  of the  $\chi$ 's. We can express  $B$  as a linear function of the  $\chi$ 's with the help of equations (22); thus

$$B = b_1 \chi_1 + b_2 \chi_2 + \dots + b_m \chi_m. \quad (23)$$

Similarly, we can express each of the quantities  $B^2, B^3, \dots, B^m$  as a linear function of the  $\chi$ 's. From the  $m$  equations thus obtained, together with the equation  $\chi(P_1) = 1$ , we can eliminate the  $m$  unknowns  $\chi_1, \chi_2, \dots, \chi_m$ , obtaining as result an algebraic equation of degree  $m$  for  $B$ ,

$$B^m + c_1 B^{m-1} + c_2 B^{m-2} + \dots + c_m = 0.$$

The  $m$  solutions of this equation give the  $m$  possible eigenvalues for  $B$ , each of which will, according to (23), be a linear function of  $b_1, b_2, \dots, b_m$  whose coefficients are a permissible set of values  $\chi'_1, \chi'_2, \dots, \chi'_m$ . These sets of values  $\chi'$  thus obtained must be all different, since if there were fewer than  $m$  different permissible sets of values  $\chi'$  for the  $\chi$ 's, there would exist a linear function of the  $\chi$ 's every one of whose eigenvalues vanishes, which would mean that the linear function itself vanishes and the  $\chi$ 's are not linearly independent. Thus the number of permissible sets of numerical values for the  $\chi$ 's is just equal to  $m$ , which is the number of classes of permutations or the number of partitions of  $n$ . This number is therefore the number of exclusive sets of states.

The properties of the  $P$ 's which are not properties of the  $\chi$ 's will only describe the degeneracy of the states, in the case of a system whose Hamiltonian does not involve the time explicitly. If  $\psi$  represents any stationary state,  $f(P)\psi$ , where  $f(P)$  is any function of the permutations, will represent another stationary state belonging to



the same energy-level, except when it vanishes identically. By expanding  $f(P)\psi$  in terms of a complete set of independent stationary states belonging to this energy-level, we get a representation of  $f(P)$  and thus of each  $P$ . In this way we see that, if we obtain a matrix representation of all the  $P$ 's consistent with each of the  $\chi$ 's being a certain number  $\chi'$ , then the number of rows and columns of the matrices will be the degree of degeneracy of the states in the exclusive set  $\chi'$ , i.e. the number of independent states belonging to each energy-level. This degeneracy is an *essential* one and cannot be removed by any perturbation that is symmetrical between all the similar particles. The states  $\psi$  and  $f(P)\psi$  are observationally indistinguishable, since any observation that can actually be made must consist in measuring an observable that is symmetrical between the similar particles and therefore commutes with  $f(P)$ . This remark applies also when the Hamiltonian involves the time explicitly.

## 60. Determination of the Energy-levels

Let us apply the perturbation method of § 46 and make a first-order calculation of the energy-levels in the case when the Hamiltonian does not involve the time explicitly. We suppose that for our unperturbed states each of the similar particles has its own 'orbit', represented by a wave function  $(q'|\alpha)$  involving only the coordinates  $q'$  of this one particle. We shall have altogether  $n$  orbits, one for each particle, which we assume for the present to be all different, and label  $\alpha_1, \alpha_2, \dots, \alpha_n$ . The wave function representing an unperturbed state of the whole system will then be the product (10). If we apply an arbitrary permutation  $P_a$  to the  $\alpha$ 's, we shall obtain another wave function

$$(q'_1|\alpha_r)(q'_2|\alpha_s)\dots(q'_n|\alpha_t) = (q'|P_a\alpha), \quad (24)$$

representing another unperturbed state with the same energy. There are thus altogether  $n!$  unperturbed states with this energy, if we assume there are no other causes of degeneracy. According to the method of § 46 when the unperturbed system is degenerate, we must consider those elements of the matrix representing the perturbing energy  $V$  that refer to two states with the same energy, i.e. those of the type  $(P_a\alpha|V|P_b\alpha)$  where  $P_a$  and  $P_b$  are two permutations of the  $\alpha$ 's. These will form a matrix with  $n!$  rows and columns, whose eigenvalues are the first-order corrections in the energy-levels.



It is necessary in the present discussion to distinguish between the two kinds of permutations, those of the  $q$ 's and those of the  $\alpha$ 's. The essential difference between them can perhaps be seen most clearly in the following way. Let us consider a permutation in the general sense, say that consisting of the interchange of 2 and 3. This may be interpreted either as the interchange of the objects 2 and 3 or as the interchange of the objects in the places 2 and 3, these two operations producing in general quite different results. The first of these interpretations is the one we have been using up to the present, the objects concerned being the  $q$ 's in the representative of a state. A permutation with this interpretation can be applied to an arbitrary function of the  $q$ 's. A permutation with the second interpretation has a meaning, however, when applied to a function of the  $q$ 's only if each of the  $q$ 's has a definite specifiable place in the function. This is not the case for a general function of the  $q$ 's, but it is the case for any of the  $n!$  functions of the type (24), the place of each  $q$  being specified by the  $\alpha$  with which it is bracketed. Any permutation applied to the  $q$ 's in given places now produces the same result as the reciprocal permutation applied to the  $\alpha$ 's. A permutation of the  $q$ 's (i.e. one with the first interpretation), since it can be applied to any function of the  $q$ 's, i.e. to the representative of any state, may be regarded as an ordinary dynamical variable. On the other hand, a permutation of places or of the  $\alpha$ 's can be considered as a dynamical variable only in a very restricted sense, since it has a meaning only when multiplied into a state whose representative is one of the  $n!$  wave functions (24) or some linear combination of them. We denote such a permutation of the  $\alpha$ 's, considered as a dynamical variable in this restricted sense, by the symbol  $P^\alpha$ .

We can form algebraic functions of the dynamical variables  $P^\alpha$  which will be other dynamical variables in the same restricted sense. In particular we can form  $\chi(P_a^\alpha)$ , the average of all  $P^\alpha$ 's similar to  $P_a^\alpha$ . This must equal  $\chi(P_a)$ , the average of the similar permutations of the  $q$ 's, since the total set of all permutations of a given type must evidently be the same whether the permutations are applied to the objects  $q$  or to the places  $\alpha$ .

If we set up arbitrarily a one-one correspondence between the  $q$ 's and the  $\alpha$ 's, as is done automatically when we label both the  $q$ 's and the  $\alpha$ 's by the numbers 1, 2, 3, ...,  $n$ , as in (10), then, if we have any permutation of the  $q$ 's, we can give a meaning to this same permu-



tion of the  $\alpha$ 's. This meaning is such that

$$(q|\alpha) = (Pq|P\alpha).$$

In this equation we can apply a permutation  $P_a$  to the  $\alpha$ 's on both sides, which will give us

$$(q|P_a\alpha) = (Pq|P_aP\alpha), \quad (25)$$

an equation which shows us the connexion between permutations of the  $q$ 's and those of the  $\alpha$ 's when applied to the wave function (24).

The matrix  $(P_a\alpha|V|P_b\alpha)$  which we must now study, may be obtained from the matrix  $(q'|V|q'')$  representing  $V$  by a coordinate transformation, in which the transformation functions are just  $(q'|P_a\alpha)$ , the wave function (24), and its conjugate complex  $(P_a\alpha|q')$ , provided these functions are properly normalized. Thus

$$(P_a\alpha|V|P_b\alpha) = \iint (P_a\alpha|q') dq' (q'|V|q'') dq'' (q''|P_b\alpha). \quad (26)$$

Again, for arbitrary  $P$ ,

$$\begin{aligned} (P_aP\alpha|V|P_bP\alpha) &= \iint (P_aP\alpha|q') dq' (q'|V|q'') dq'' (q''|P_bP\alpha) \\ &= \iint (P_aP\alpha|Pq') dq' (Pq'|V|Pq'') dq'' (Pq''|P_bP\alpha), \end{aligned}$$

when we apply the permutation  $P$  to the variables of integration  $q'$  and  $q''$ . With the help of (25), this reduces to

$$(P_aP\alpha|V|P_bP\alpha) = \iint (P_a\alpha|q') dq' (Pq'|V|Pq'') dq'' (q''|P_b\alpha). \quad (27)$$

Now since  $V$  is symmetrical between all the particles, we must have

$$(q'|V|q'') = (Pq'|V|Pq''),$$

like (1), and hence, comparing (26) and (27), we obtain

$$(P_a\alpha|V|P_b\alpha) = (P_aP\alpha|V|P_bP\alpha). \quad (28)$$

Let  $(P\alpha|V|\alpha) = V_P$  for brevity. Then, taking  $P = P_b^{-1}$  in (28), we obtain

$$(P_a\alpha|V|P_b\alpha) = (P_aP_b^{-1}\alpha|V|\alpha) = V_{P_aP_b^{-1}}.$$

Thus the general matrix element  $(P_a\alpha|V|P_b\alpha)$  depends only on the ratio  $P_aP_b^{-1}$ , and of the total of  $(n!)^2$  matrix elements there are only  $n!$  different ones. The coefficient of any  $V_P$  in the matrix will be a matrix, each of whose elements is 0 or 1, the 1 occurring when

$$(P_a\alpha|V|P_b\alpha) = V_P,$$

i.e. when  $P_aP_b^{-1} = P$ . But the latter matrix, multiplied into any wave function  $(q|P_b\alpha)$ , gives the result  $(q|P_a\alpha)$  with  $P_aP_b^{-1} = P$ , i.e. it gives



the result  $(q|PP_b\alpha)$ , so that it is precisely the matrix representing the dynamical variable  $P^\alpha$  or the permutation  $P$  applied to the  $\alpha$ 's. Thus the whole matrix  $(P_a\alpha|V|P_b\alpha)$  is equal to the matrix representing  $\sum_P V_P P^\alpha$ , where the summation is over all the  $n!$  permutations  $P$ , and we can put

$$V = \sum_P V_P P^\alpha. \quad (29)$$

This formula shows that the perturbing energy  $V$  is equal to a linear function of the permutation dynamical variables  $P^\alpha$  with numerical coefficients  $V_P$ . It is, of course, only an approximate formula, as it holds only with neglect of those matrix elements of  $V$  that refer to two different energy-levels of the unperturbed system. It can, however, be used for the calculation of the energy-levels in the first approximation, and is very convenient for this purpose as the expression  $\sum_P V_P P^\alpha$  is easily handled. This expression, it should be remembered, is a dynamical variable only in the restricted sense mentioned above, but this sense is sufficiently general for equation (29) to be valid with neglect of those matrix elements of  $V$  referring to two different energy-levels of the unperturbed system.

As an example of an application of (29) we shall determine the average energy of all those states, arising from a given state of the unperturbed system, that belong to one exclusive set. This requires us to calculate the average eigenvalue of  $V$  when the  $\chi$ 's have specified numerical values  $\chi'$ . Now the average eigenvalue of  $P_a^\alpha$  equals that of  $P^\alpha P_a^\alpha (P^\alpha)^{-1}$  for arbitrary  $P^\alpha$  and thus equals that of  $n!^{-1} \sum_{P^\alpha} P^\alpha P_a^\alpha (P^\alpha)^{-1}$ , which is  $\chi'(P_a^\alpha)$  or  $\chi'(P_a)$ . Hence the average eigenvalue of  $V$  is  $\sum_P V_P \chi'(P)$ . A similar method could be used for calculating the average eigenvalue of any function of  $V$ , it being necessary only to replace each  $P^\alpha$  by  $\chi(P)$  to perform the averaging.

The number of energy-levels in an exclusive set  $\chi = \chi'$  that arise from a given state of the unperturbed system is equal to the number of eigenvalues of (29) that are consistent with the equations  $\chi = \chi'$ . This number is the number of rows and columns in a representation of the  $P$ 's in which each  $\chi = \chi'$ , which number, from the result at the end of the preceding section, is just the degree of degeneracy of the states in this set.

The modifications required in the theory when the orbits  $\alpha_1, \alpha_2, \dots, \alpha_n$  of the undisturbed system are not all different may easily be made.



Suppose, for example, that  $\alpha_1$  and  $\alpha_2$  are the same. Then the permutation  $P_{12}^\alpha$  that causes an interchange of  $\alpha_1$  and  $\alpha_2$  must equal unity. Only functions of the  $P^\alpha$ 's that commute with  $P_{12}^\alpha$  now have a meaning. This, however, is sufficient for us to be able to follow out the same sort of argument as before and obtain a result of the same form (29). The term in the summation in (29) that involves the permutation  $P_{12}^\alpha$  now does not occur, since it could be added on to the term involving the identical permutation  $P_1^\alpha$ . For the remaining terms, any two terms  $P_a^\alpha$  and  $P_b^\alpha$  must have the same coefficient if the permutations  $P_a^\alpha$  and  $P_b^\alpha$  can be obtained from one another by the interchange of  $\alpha_1$  and  $\alpha_2$ . This results in  $\sum_P V_P P^\alpha$  commuting with  $P_{12}^\alpha$  and thus having a meaning. The condition  $P_{12}^\alpha = 1$  imposes restrictions on the possible numerical values  $\chi'$  that the  $\chi$ 's can have and reduces the number of characters.

## 61. Application to Electrons

Let us now consider the case when the similar particles are electrons. This requires, according to Pauli's exclusion principle discussed in § 57, that we take into account only the antisymmetrical states. It is now necessary to make explicit reference to the fact that electrons have spins, which show themselves through an angular momentum and a magnetic moment. The effect of the spin on the motion of an electron in an electromagnetic field is not very great. There are additional forces on the electron due to its magnetic moment, requiring additional terms in the Hamiltonian. The spin angular momentum does not have any direct action on the motion, but it comes into play when there are forces tending to rotate the magnetic moment, since the magnetic moment and angular momentum are constrained to be always in the same direction. These effects are all small, however, of the same order of magnitude as that of the relativistic variation of mass with velocity, so there would be no point in taking them into account in a non-relativistic theory. The importance of the spin lies not in these small effects on the motion of the electron, but in the fact that it gives two internal states to the electron, corresponding to the two possible values of the spin component in any assigned direction, which causes a doubling in the number of independent states of an electron moving in a given field. This fact has far-reaching consequences when combined with Pauli's exclusion principle.

For the complete description of an electron we require the spin



dynamical variables  $\sigma$ , which were introduced in § 19 and whose connexion with the spin angular momentum was given in § 39, together with the Cartesian coordinates  $x, y, z$  and momenta  $p_x, p_y, p_z$ . The spin dynamical variables are assumed to commute with these coordinates and momenta. Thus a complete set of commuting observables for a system consisting of a single electron will be  $x, y, z, \sigma_z$ . In a representation in which these are diagonal, the representative of any state will be a function of four variables  $x', y', z', \sigma'_z$ . Since  $\sigma'_z$  has a domain consisting of only two points, namely 1 and  $-1$ , this function of four variables is the same as two functions of three variables, namely the two functions

$$(x'y'z'|)_+ = (x', y', z', +1|) \quad (x'y'z'|)_- = (x', y', z', -1|).$$

Thus the presence of the spin may be considered either as introducing a new variable into the representative of a state or as giving this representative two components.

In our present work on the theory of several electrons, we shall consider the spins as giving extra variables in the representatives of states. For brevity, we shall write the single variable  $\mathbf{x}_r$  instead of  $x_r, y_r, z_r$  for the coordinates of the  $r$ -th electron and shall omit the suffix  $z$  from  $\sigma_{zr}$  when it occurs in representatives. Thus the representative of a state when there are  $n$  electrons will be written

$$(\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_n \sigma_1 \sigma_2 \dots \sigma_n|) = (\mathbf{x}, \sigma|), \quad (30)$$

for brevity. The exclusion principle requires that (30) shall be anti-symmetrical in the  $\mathbf{x}$ 's and  $\sigma$ 's together, i.e. if any permutation is applied to the  $\mathbf{x}$ 's and also to the  $\sigma$ 's, (30) must remain unchanged or change sign according to whether the permutation is even or odd. In symbols

$$(\mathbf{x}, \sigma|) = \pm (P\mathbf{x}, P\sigma|) \quad (31)$$

for any permutation  $P$ . Thus even if we neglect the spin forces in the Hamiltonian, we must take the spin variables into account in order to determine what states are allowed by the exclusion principle.

If the theory of the three preceding sections is applied directly to the case of electrons, it will not give anything of interest, since all the allowed states are eigenstates of any permutation belonging to the eigenvalue  $\pm 1$ . We may, however, consider permutations  $P$  which operate on the  $\mathbf{x}$ -variables alone in the representative of a state, and apply our theory to these. Such permutations may also be considered as dynamical variables. Further, they are also constants of the motion



when we neglect the terms in the Hamiltonian that arise from the spin forces, since this neglect results in the Hamiltonian not involving the spin dynamical variables  $\sigma$  at all. Hence with these permutations  $P$  we can again introduce the  $\chi$ 's, equal to the average of all of the  $P$ 's in each class, and assert that for any permissible set of numerical values  $\chi'$  for the  $\chi$ 's there will be one exclusive set of states. Thus there exist these exclusive sets of states for systems containing many electrons even when we restrict ourselves to a consideration of only those states that satisfy Pauli's principle. The exclusiveness of the sets of states is now, of course, only approximate, since the  $\chi$ 's are constants only so long as we neglect the spin forces. There will actually be a small probability for a transition from a state in one set to a state in another.

From (31) we obtain  $PP^\sigma = \pm 1$ , (32)

where  $P$  denotes any permutation which operates on the  $x$ -variables and  $P^\sigma$  the same permutation operating on the  $\sigma$ -variables in the representative of a state. There is thus a simple connexion between the  $P$ 's and  $P^\sigma$ 's, which means that instead of studying the dynamical variables  $P$  we can get all the results we want, e.g. the characters  $\chi'$ , by studying the dynamical variables  $P^\sigma$ . The  $P^\sigma$ 's are much easier to study on account of the fact that the  $\sigma$  variables in the wave function have domains consisting each of only the two points 1 and  $-1$ , which are the two eigenvalues of each  $\sigma_z$ . This fact results in there being fewer characters  $\chi'$  for the group of permutations of the  $\sigma$ -variables than for the group of general permutations, since it prevents a function of the variables  $\sigma_1, \sigma_2, \dots, \sigma_n$  from being anti-symmetrical in more than two of them.

The study of the dynamical variables  $P^\sigma$  is made specially easy by the fact that we can express them as algebraic functions of the dynamical variables  $\sigma$ . Consider the quantity

$$O_{12} = \frac{1}{2}\{1 + (\sigma_1, \sigma_2)\}.$$

With the help of equations (54) and (55) of § 19 we find readily that

$$(\sigma_1, \sigma_2)^2 = (\sigma_{x1}\sigma_{x2} + \sigma_{y1}\sigma_{y2} + \sigma_{z1}\sigma_{z2})^2 = 3 - 2(\sigma_1, \sigma_2), \quad (33)$$

and hence that

$$O_{12}^2 = \frac{1}{4}\{1 + 2(\sigma_1, \sigma_2) + (\sigma_1, \sigma_2)^2\} = 1. \quad (34)$$

Again, we find

$$\begin{aligned} O_{12}\sigma_{x1} &= \frac{1}{2}\{\sigma_{x1} + \sigma_{x2} - i\sigma_{z1}\sigma_{y2} + i\sigma_{y1}\sigma_{z2}\} \\ \sigma_{x2}O_{12} &= \frac{1}{2}\{\sigma_{x2} + \sigma_{x1} + i\sigma_{y1}\sigma_{z2} - i\sigma_{z1}\sigma_{y2}\} \end{aligned}$$



and hence

$$O_{12} \sigma_{x1} = \sigma_{x2} O_{12}.$$

Similar relations hold for  $\sigma_{y1}$  and  $\sigma_{z1}$  so that we have

$$O_{12} \sigma_1 = \sigma_2 O_{12}$$

or

$$O_{12} \sigma_1 O_{12}^{-1} = \sigma_2.$$

From this we can obtain with the help of (34)

$$O_{12} \sigma_2 O_{12}^{-1} = \sigma_1.$$

These commutability relations for  $O_{12}$  with  $\sigma_1$  and  $\sigma_2$  are precisely the same as those for  $P_{12}^\sigma$ , the permutation consisting of the interchange of the spin variables of electrons 1 and 2. Thus we can put

$$O_{12} = c P_{12}^\sigma,$$

where  $c$  is a number. Equation (34) shows that  $c = \pm 1$ . To determine which of these values for  $c$  is the correct one, we observe that the eigenvalues of  $P_{12}^\sigma$  are 1, 1, 1, -1, corresponding to the fact that there exist three independent symmetrical and one antisymmetrical function of the two variables  $\sigma_{z1}, \sigma_{z2}$ , namely, with the notation of § 19, the three symmetrical functions  $f_\alpha(\sigma_1)f_\alpha(\sigma_2)$ ,  $f_\beta(\sigma_1)f_\beta(\sigma_2)$ ,  $f_\alpha(\sigma_1)f_\beta(\sigma_2) + f_\beta(\sigma_1)f_\alpha(\sigma_2)$ , and the one antisymmetrical function  $f_\alpha(\sigma_1)f_\beta(\sigma_2) - f_\beta(\sigma_1)f_\alpha(\sigma_2)$ . Thus the mean of the eigenvalues of  $P_{12}^\sigma$  is  $\frac{1}{2}$ . Now the mean of the eigenvalues of  $(\sigma_1, \sigma_2)$  is evidently zero and hence the mean of the eigenvalues of  $O_{12}$  is  $\frac{1}{2}$ . Thus we must have  $c = +1$ , and so we can put

$$P_{12}^\sigma = \frac{1}{2}\{1 + (\sigma_1, \sigma_2)\}.$$

In this way any permutation  $P^\sigma$  consisting simply of an interchange can be expressed as an algebraic function of the  $\sigma$ 's. Any other permutation  $P^\sigma$  can be expressed as a product of interchanges and can therefore also be expressed as a function of the  $\sigma$ 's. With the help of (32) we can now express the  $P$ 's as algebraic functions of the  $\sigma$ 's and eliminate the  $P^\sigma$ 's from the discussion. We have, since the — sign must be taken in (32) when the permutations are interchanges and since the square of an interchange is unity,

$$P_{12} = -\frac{1}{2}\{1 + (\sigma_1, \sigma_2)\}. \quad (35)$$

The formula (35) may conveniently be used for the evaluation of the characters  $\chi'$  which define the exclusive sets of states. We have, for example, for the permutations consisting of interchanges

$$\chi_{12} = \chi(P_{12}) = -\frac{1}{2}\left\{1 + \frac{2}{n(n-1)} \sum_{r < t} (\sigma_r, \sigma_t)\right\}.$$



If we introduce the dynamical variable  $s$  to describe the magnitude of the total spin angular momentum,  $\frac{1}{2} \sum_r \sigma_r$  in units of  $\hbar$ , through the formula

$$s(s+1) = \left( \frac{1}{2} \sum_r \sigma_r, \frac{1}{2} \sum_t \sigma_t \right),$$

analogous to equation (12) of Chapter VII, we have

$$\begin{aligned} 2 \sum_{r < t} (\sigma_r, \sigma_t) &= \left( \sum_r \sigma_r, \sum_t \sigma_t \right) - \sum_r (\sigma_r, \sigma_r) \\ &= 4s(s+1) - 3n. \end{aligned}$$

Hence

$$\chi_{12} = -\frac{1}{2} \left\{ 1 + \frac{4s(s+1) - 3n}{n(n-1)} \right\} = -\frac{n(n-4) + 4s(s+1)}{2n(n-1)}. \quad (36)$$

Thus  $\chi_{12}$  is expressible as a function of the dynamical variable  $s$  and of  $n$  the number of electrons. Any of the other  $\chi$ 's could be evaluated on similar lines and would have to be a function of  $s$  and  $n$  only, since there are no other symmetrical functions of all the  $\sigma$  dynamical variables which could be involved. There is therefore one set of numerical values  $\chi'$  for the  $\chi$ 's, and thus one exclusive set of states, for each eigenvalue  $s'$  of  $s$ . The eigenvalues of  $s$  are

$$\frac{1}{2}n, \frac{1}{2}n-1, \frac{1}{2}n-2, \dots$$

the series terminating with 0 or  $\frac{1}{2}$ .

We see in this way that each of the stationary states of a system with several electrons is an eigenstate of  $s$ , the magnitude in units of  $\hbar$  of the total spin angular momentum  $\frac{1}{2} \sum_r \sigma_r$ , belonging to a definite eigenvalue  $s'$ . For any given  $s'$  there will be  $2s'+1$  possible values for a component of the total spin vector in any direction and these will correspond to  $2s'+1$  independent stationary states with the same energy. When we do not neglect the forces due to the spin magnetic moments these  $2s'+1$  states will in general be split up into  $2s'+1$  states with slightly different energies, and will thus form a multiplet of multiplicity  $2s'+1$ . Transitions in which  $s'$  changes, i.e. transitions from one multiplicity to another, cannot occur when the spin forces are neglected and will have only a small probability of occurrence when the spin forces are not neglected.

We can determine the energy-levels of a system with several electrons to the first approximation by using formula (29). If we consider only the Coulomb forces between the electrons, then the interaction energy  $V$  will consist of a sum of parts each referring to



only two electrons, which will result in all the matrix elements  $V_P$  vanishing except those for which  $P$  is the identical permutation or is simply an interchange of two electrons. Thus (29) will reduce to

$$V = V_1 + \sum_{r < s} V_{rs} P_{rs}^{\alpha}, \quad (37)$$

$V_{rs}$  being the matrix element referring to the interchange of orbits  $r$  and  $s$ . Since the  $P^{\alpha}$ 's have the same properties as the  $P$ 's, any function of the  $P^{\alpha}$ 's will have the same eigenvalues as the corresponding function of the  $P$ 's, so that the right-hand side of (37) will have the same eigenvalues as

$$V_1 + \sum_{r < s} V_{rs} P_{rs}$$

or

$$V_1 - \frac{1}{2} \sum_{r < s} V_{rs} \{1 + (\sigma_r, \sigma_s)\} \quad (38)$$

from (35). The eigenvalues of (38) will give the first-order corrections in the energy-levels. The form of (38) shows that a model which assumes a coupling energy between the spins of the various electrons, of magnitude  $-\frac{1}{2}V_{rs}(\sigma_r, \sigma_s)$  for the electrons in the  $r$  and  $s$  orbits, would meet with a fair amount of success. This coupling energy is much greater than that of the spin magnetic moments. Such models of the atom were in use before the justification by quantum mechanics was obtained.

If two of the orbits of our unperturbed system are the same, say the orbits  $\alpha_1$  and  $\alpha_2$  are the same, we must take only those eigenvalues of (37) that are consistent with  $P_{12}^{\alpha} = 1$ , or those eigenvalues of (38) consistent with  $P_{12} = 1$  or  $P_{12}^{\sigma} = -1$ . This means we must take only those eigenvalues of (38) belonging to eigenfunctions that are simultaneously eigenfunctions of  $P_{12}^{\sigma}$  belonging to the eigenvalue  $-1$ , i.e. eigenfunctions that are antisymmetrical in  $\sigma_1$  and  $\sigma_2$ . Thus we may say that the two electrons in the orbits  $\alpha_1$  and  $\alpha_2$  have their spins antiparallel. The case of more than two orbits the same cannot occur with electrons.



# THEORY OF RADIATION

## 62. Second Quantization

WE shall begin this chapter by considering some general properties of an assembly of  $n$  similar systems of any kind that satisfy the Einstein-Bose statistics. If we take a representation in which sets of observables  $q_1, q_2, \dots, q_n$  describing the first, second, ..., last system respectively, are diagonal, the representative  $(q'_1 q'_2 \dots q'_n |)$  of any state must be symmetrical in the variables  $q'_1, q'_2, \dots, q'_n$ . Suppose the eigenvalues of any of the  $q$ 's,  $q_r$  say, are  $q^{(1)}, q^{(2)}, q^{(3)}, \dots$ , which we assume for definiteness to be discrete. These eigenvalues must be the same for each of the  $n$  systems, i.e. they must be independent of  $r$ . (They will each be in general a set of numbers, consisting of an eigenvalue of each of the set of commuting observables  $q_r$ .) If we now have any *symmetrical* function of the variables  $q'_1, q'_2, \dots, q'_n$ , each point in the domain of this function can be specified by  $n'_1, n'_2, n'_3, \dots$ , the numbers of  $q$ 's equal to  $q^{(1)}, q^{(2)}, q^{(3)}, \dots$  respectively. The variables  $n'_1, n'_2, n'_3, \dots$  will do just as well as the variables  $q'_1, q'_2, \dots, q'_n$ , so long as we are dealing only with symmetrical functions. Thus the representatives of states of our assembly satisfying the Einstein-Bose statistics may be expressed as functions of the variables  $n'_1, n'_2, n'_3, \dots$  instead of the variables  $q'_1, q'_2, \dots, q'_n$ . This change is effectively a transformation to a new representation in which the rows and columns of the matrices are labelled by the observables  $n_1, n_2, n_3, \dots$ , which observables are the numbers of systems with  $q$ 's equal to  $q^{(1)}, q^{(2)}, q^{(3)}, \dots$  respectively, or, as we may say, the numbers of systems in the states  $q^{(1)}, q^{(2)}, q^{(3)}, \dots$ .

Since the new observables  $n_1, n_2, n_3, \dots$  are functions of the  $q_1, q_2, \dots, q_n$  (non-analytic functions, it is true) the transformation is of the trivial kind consisting essentially of a relabelling of the rows and columns, and the only change to be made in the representative of a state will be that arising from the change in the weights of the different points of its domain. To determine this change we use the condition

$$\sum_{n_1 n_2 \dots} |(n_1 n_2 \dots |)|^2 = \sum_{q_1 q_2 \dots q_n} |(q_1 q_2 \dots q_n |)|^2,$$

from which we can infer that

$$|(n_1 n_2 \dots |)|^2 = \sum |(q_1 q_2 \dots q_n |)|^2, \quad (1)$$



the summation in (1) being over all values of the  $q$ 's such that  $n_1$  of them are equal to  $q^{(1)}$ ,  $n_2$  equal to  $q^{(2)}$ , and so on. The number of terms in the summation in (1) is  $n!/(n_1!n_2!n_3!\dots)$  and they are all equal, on account of  $(q_1 q_2 \dots q_n)$  being symmetrical. It is thus clear that we must take

$$(n_1 n_2 \dots |) = [n!/n_1!n_2!n_3!\dots]^{\frac{1}{2}}(q_1 q_2 \dots q_n |). \quad (2)$$

We must now obtain the transformation law for the representatives of dynamical variables from the  $q$ -representation to the  $n$ -representation. As this problem is rather complicated for a general dynamical variable, we shall here deal only with the special case† when the dynamical variable is of the form

$$U = \sum_r U_r, \quad (3)$$

$U_r$  being a function only of the variables describing the  $r$ -th system and the form of  $U_r$  in terms of these variables being the same for all  $r$ , so as to make  $U$  symmetrical between all the systems, as it must be if it is to have any physical significance. The representative of  $U_r$  in the  $q_r$ -representation will be  $(q'_r|U_r|q''_r)$ , which will be a matrix independent of  $r$ , i.e. the same for each of the  $n$  systems. Its elements may also be written  $(q^{(a)}|U|q^{(b)})$  or  $U_{ab}$  for brevity. The representative of  $U$  in the complete  $q$ -representation will thus be

$$\begin{aligned} (q'_1 q'_2 \dots q'_n | U | q''_1 q''_2 \dots q''_n) \\ = \sum_r (q'_r | U_r | q''_r) \delta_{q'_1 q''_1} \delta_{q'_2 q''_2} \dots \delta_{q'_{r-1} q''_{r-1}} \delta_{q'_{r+1} q''_{r+1}} \dots \delta_{q'_n q''_n}. \end{aligned} \quad (4)$$

A convenient way of transforming this representative to the  $n$ -representation is to take the equation

$$\psi_2 = U\psi_1 \quad (5)$$

and transform the representative of this equation. From (4), this equation will be represented in the  $q$ -representation by

$$\begin{aligned} (q_1 q_2 \dots q_n | 2) = \sum_r (q_r | U_r | q_r) (q_1 q_2 \dots q_n | 1) + \\ + \sum_r \sum_{q'_r \neq q_r} (q_r | U_r | q'_r) (q_1 q_2 \dots q_{r-1} q'_r q_{r+1} \dots q_n | 1), \end{aligned} \quad (6)$$

the terms arising from the diagonal matrix elements of  $U$  being separated from the non-diagonal ones for convenience later. If we now make the transformation to the  $n$ -representation, using equation

† The general case has been dealt with by Jordan, *Z. f. Physik*, **45** (1927), 774.



(2), equation (6) becomes

$$(n_1 n_2 \dots | 2) = \sum_r (q_r | U_r | q_r) (n_1 n_2 \dots | 1) + \\ + \sum_r \sum_{q'_r \neq q_r} [(n_{q'_r} + 1)/n_{q_r}]^{\frac{1}{2}} (q_r | U_r | q'_r) (n_1 n_2 \dots n_{q_r} - 1 \dots n_{q'_r} + 1 \dots | 1), \quad (7)$$

after removal of the factor  $[n_1! n_2! n_3! \dots / n!]^{\frac{1}{2}}$  throughout. The sum  $\sum_r (q_r | U_r | q_r)$  in (7) means a sum of terms each of the type  $(q^{(a)} | U | q^{(a)})$  or  $U_{aa}$ , the number of times this typical term occurs being the number of  $q$ 's that equal  $q^{(a)}$ , which is just  $n_a$ . Thus this sum is equal to  $\sum_a n_a U_{aa}$ .

Again, the double sum  $\sum_r \sum_{q'_r \neq q_r}$  in (7) consists of terms each of the type  $[(n_b + 1)/n_a]^{\frac{1}{2}} U_{ab} (n_1 n_2 \dots n_a - 1 \dots n_b + 1 \dots | 1)$  with  $b \neq a$ . The number of times this typical term occurs is equal to the number of ways of choosing  $r$  and  $q'_r$  such that  $q_r = q^{(a)}$  and  $q'_r = q^{(b)}$ . This is just  $n_a$ , the number of ways of choosing  $r$  such that  $q_r = q^{(a)}$ , since there is always just one way of choosing  $q'_r = q^{(b)}$ . Equation (7) thus reduces to

$$(n_1 n_2 \dots | 2) = \sum_a n_a U_{aa} (n_1 n_2 \dots | 1) + \\ + \sum_a \sum_{b \neq a} n_a^{\frac{1}{2}} (n_b + 1)^{\frac{1}{2}} U_{ab} (n_1 n_2 \dots n_a - 1 \dots n_b + 1 \dots | 1), \quad (8)$$

which may be written

$$(n_1 n_2 \dots | 2) = \sum_{ab} n_a^{\frac{1}{2}} (n_b + 1 - \delta_{ab})^{\frac{1}{2}} U_{ab} (n_1 n_2 \dots n_a - 1 \dots n_b + 1 \dots | 1), \quad (9)$$

if by  $(n_1 n_2 \dots n_a - 1 \dots n_b + 1 \dots | 1)$  when  $b = a$  we understand simply  $(n_1 n_2 \dots n_a \dots | 1)$ .

The eigenvalues of each of our new dynamical variables  $n_1, n_2, \dots$  are the integers 0, 1, 2, 3, ... They are thus the same, apart from the factor  $\hbar$ , as those of the action variable  $J$  in the problem of the simple harmonic oscillator, when the arbitrary additive constant in this action variable is chosen as in equation (57) of § 36. Hence each  $n_a$  is a dynamical variable of the same nature as the action variable of a simple harmonic oscillator and we can introduce an angle variable  $w_a$  canonically conjugate to it, or rather we can introduce  $e^{iw_a}$  and  $e^{-iw_a}$ . Corresponding to equations (59) of § 36 we shall have

$$\left. \begin{aligned} e^{iw_a} n_a &= (n_a - 1) e^{iw_a} \\ e^{-iw_a} n_a &= (n_a + 1) e^{-iw_a} \end{aligned} \right\} \quad (10)$$

Also we have that  $e^{iw_a}$ ,  $e^{-iw_a}$  and  $n_a$  commute with  $e^{iw_b}$ ,  $e^{-iw_b}$  and  $n_b$  for  $b \neq a$ .



The new dynamical variables  $e^{iw_a}$ ,  $e^{-iw_a}$  are defined by their matrix representatives in a representation in which  $n_a$  is diagonal, like the  $e^{iw}$ ,  $e^{-iw}$  of § 36. From the form of these matrix representatives it follows that when  $e^{-iw_a}$  is multiplied into a  $\psi$ -symbol whose representative is  $(n_1 n_2 \dots n_a \dots |)$ , the representative of the product is

$$(n_1 n_2 \dots n_a + 1 \dots |),$$

and when  $e^{iw_a}$  is multiplied into this  $\psi$ -symbol, the representative of the product is

$$\begin{aligned} (n_1 n_2 \dots n_a - 1 \dots |) & \quad \text{for } n_a \geq 1 \\ 0 & \quad \text{for } n_a = 0. \end{aligned}$$

This means that when  $e^{-iw_a}$  and  $e^{iw_a}$  are multiplied into  $\psi$ -symbols, they are equivalent to the operations of substitution of  $n_a + 1$  and  $n_a - 1$  for  $n_a$  respectively, the second substitution being understood to give the result zero for  $n_a = 0$ . We can now see that equation (9) is just the representative of

$$\psi_2 = \sum_{ab} n_a^{\frac{1}{2}} (n_b + 1 - \delta_{ab})^{\frac{1}{2}} U_{ab} e^{iw_a} e^{-iw_b} \psi_1. \quad (11)$$

Equation (11) must hold whenever (5) holds and hence

$$\begin{aligned} U &= \sum_{ab} n_a^{\frac{1}{2}} (n_b + 1 - \delta_{ab})^{\frac{1}{2}} U_{ab} e^{iw_a} e^{-iw_b} \\ &= \sum_{ab} n_a^{\frac{1}{2}} e^{iw_a} U_{ab} (n_b + 1)^{\frac{1}{2}} e^{-iw_b}, \end{aligned} \quad (12)$$

with the help of (10). This gives us  $U$  in terms of the new variables  $n$  and their conjugates, and provides us immediately with the representative of  $U$  in the new representation. The  $U_{ab}$  here are, of course, just numerical coefficients.

We can put the result (12) into a simpler form by introducing the dynamical variables

$$\left. \begin{aligned} \xi_a &= (n_a + 1)^{\frac{1}{2}} e^{-iw_a} = e^{-iw_a} n_a^{\frac{1}{2}} \\ \bar{\xi}_a &= e^{iw_a} (n_a + 1)^{\frac{1}{2}} = n_a^{\frac{1}{2}} e^{iw_a} \end{aligned} \right\} \text{and their conjugate complexes} \quad (13)$$

These dynamical variables are of the form of (61) of Chapter VI and correspond, apart from numerical coefficients, to  $p - iq$  and  $p + iq$  in the problem of the simple harmonic oscillator. We have

$$\begin{aligned} \xi_a \bar{\xi}_a &= n_a + 1 \\ \bar{\xi}_a \xi_a &= n_a, \end{aligned} \quad (14)$$



and thus, since variables with different suffixes  $a$  commute, the complete set of quantum conditions for the  $\xi$ 's and  $\bar{\xi}$ 's is

$$\left. \begin{aligned} \xi_a \xi_b - \xi_b \xi_a &= 0 \\ \bar{\xi}_a \bar{\xi}_b - \bar{\xi}_b \bar{\xi}_a &= 0 \\ \xi_a \bar{\xi}_b - \bar{\xi}_b \xi_a &= \delta_{ab}. \end{aligned} \right\} \quad (15)$$

Expressed in terms of the  $\xi$ 's and  $\bar{\xi}$ 's, equation (12) takes the simple form

$$U = \sum_{ab} \bar{\xi}_a U_{ab} \xi_b. \quad (16)$$

We could carry through all the preceding work with reference to a different initial representation, say one in which the observables  $Q_1, Q_2, \dots, Q_n$  describing the first, second, ..., last system respectively are diagonal, instead of  $q_1, q_2, \dots, q_n$ . If  $Q^{(A)}, Q^{(B)}, \dots$  are the eigenvalues of a  $Q$ , we should introduce observables  $n_A, n_B, \dots$  giving the numbers of systems with  $Q$ 's equal to  $Q^{(A)}, Q^{(B)}, \dots$  respectively. Corresponding to these new  $n$ 's we should have new  $w$ 's, say  $w_A, w_B, \dots$  (defined only in exponentials like  $e^{iw_A}, e^{-iw_A}$ ), and new  $\xi$ 's and  $\bar{\xi}$ 's, say  $\xi_A, \xi_B, \dots$  and  $\bar{\xi}_A, \bar{\xi}_B, \dots$ . The new equation (16) would read

$$U = \sum_{AB} \bar{\xi}_A U_{AB} \xi_B, \quad (17)$$

where

$$\begin{aligned} U_{AB} &= (Q^{(A)}|U|Q^{(B)}) = \sum_{ab} (Q^{(A)}|q^{(a)})(q^{(a)}|U|q^{(b)})(q^{(b)}|Q^{(B)}) \\ &= \sum_{ab} (Q^{(A)}|q^{(a)})U_{ab}(q^{(b)}|Q^{(B)}), \end{aligned} \quad (18)$$

$(Q^{(A)}|q^{(a)})$  and  $(q^{(b)}|Q^{(B)})$  being the transformation functions between the  $q$ 's and  $Q$ 's for a single system of the assembly. Equating the right-hand sides of (16) and (17) and using (18), we get

$$\sum_{ab} \bar{\xi}_a U_{ab} \xi_b = \sum_{ABab} \bar{\xi}_A (Q^{(A)}|q^{(a)})U_{ab}(q^{(b)}|Q^{(B)})\xi_B. \quad (19)$$

Now  $U_r$  in (3) can be an arbitrary function of the dynamical variables describing the  $r$ -th system and so the matrix elements  $U_{ab}$  can be arbitrary. Since (19) holds with arbitrary  $U_{ab}$ , we must have

$$\bar{\xi}_a = \sum_A \bar{\xi}_A (Q^{(A)}|q^{(a)}) \quad \xi_b = \sum_B (q^{(b)}|Q^{(B)})\xi_B. \quad (20)$$

These equations give the transformation laws for the  $\xi$ 's and  $\bar{\xi}$ 's.

Equations (20) show the existence of a remarkable analogy between the variables  $\xi_a$ , which may all be regarded as forming one operator involving the parameter  $q^{(a)}$ , and the representative  $(q^{(a)}|)$  of a state of a single system of the assembly. These two functions of  $q^{(a)}$  have



precisely the same transformation law under a passage from  $q$  to  $Q$ . Further, the interpretation (14) for  $\bar{\xi}_a \xi_a$  is to some extent analogous to the ordinary physical interpretation of  $|(q^{(a)})|^2$ . The analogy extends also to equations of motion. If we suppose the systems of our Einstein-Bose assembly to be all moving under the action of some external field of force, with no interaction between the systems, the total Hamiltonian for the assembly will be of the form of  $U$  in (3), where  $U_r$  is the Hamiltonian for the  $r$ -th system alone, moving under the action of the external field of force. Taking  $U$  as Hamiltonian, we get as equations of motion for the  $\xi$ 's, from (16) and the quantum conditions (15).

$$\begin{aligned} i\hbar \dot{\xi}_a &= \xi_a U - U \xi_a \\ &= \xi_a \sum_{cb} \bar{\xi}_c U_{cb} \xi_b - \sum_{cb} \bar{\xi}_c U_{cb} \xi_b \xi_a \\ &= \sum_b U_{ab} \xi_b. \end{aligned} \tag{21}$$

This is of the same form as the Schrödinger wave equation for the  $r$ -th system alone with the Hamiltonian  $U_r$ ,  $\xi_a$  playing the part of the wave function  $(q_r^{(a)}|)$ .

We have now established the general result that the transformation equations and equations of motion for the  $\xi$ 's describing an Einstein-Bose assembly of systems acted on by an external field of force may be obtained from the corresponding equations for the wave function describing a single system of the assembly by the application of a certain definite procedure, which is called *second quantization*. This consists in assuming that the wave function  $(q^{(a)}|)$  describing the single system is not a numerical function of the parameter  $q^{(a)}$ , but is an operator for each  $q^{(a)}$ , satisfying the quantum conditions (15). It then goes over into  $\xi_a$ , the form of its wave equation and transformation law remaining unaltered.

### 63. Waves and Einstein-Bose Particles

The theory of the preceding section provides the mathematical basis for the reconciliation of the wave and corpuscular pictures of light. It shows that an assembly of particles satisfying the Einstein-Bose statistics may be described by dynamical variables  $n_a$  and their conjugates, which are formally the same as the action and angle variables describing simple harmonic oscillators. Thus an Einstein-Bose assembly is dynamically equivalent to a set of simple harmonic oscillators, there being one oscillator corresponding to each of a complete set of



*independent states of a system of the assembly, the action variable of the oscillator corresponding to the number of systems in the state.* We may replace the set of simple harmonic oscillators by a train of waves, each Fourier component of the waves being dynamically equivalent to a simple harmonic oscillator. We then see that our Einstein-Bose assembly is dynamically equivalent to a system of waves. Thus if we have any vibrating medium which we wish to deal with according to quantum mechanics, we may treat it either as a system of waves or as an assembly of Einstein-Bose particles, the two points of view being consistent and mathematically equivalent.

We may apply our theory of Einstein-Bose assemblies to the case when the  $n_a$ 's are all large, so that the  $\xi_a$ 's are also large and we may neglect the  $\delta_{ab}$  on the right-hand side of (15). With this approximation our dynamical variables  $\xi_a, \bar{\xi}_a$  all commute with each other and may be counted as numbers, and the equations of motion (21) become ordinary differential equations between numbers. These equations are now identical with the Schrödinger equation for a single one of the systems perturbed by the external field of force, the set of numbers  $\xi_a$  playing the part of the wave function  $(q^{(a)}|)$ . If this wave function is normalized to  $n$ , it may be considered to represent an assembly of  $n$  independent systems in the way discussed in § 51. The interpretation of the wave function, namely the interpretation of  $|(q^{(a)}|)^2$  as the probable number of systems in state  $q^{(a)}$ , now corresponds exactly to the interpretation of the  $\xi_a$ 's provided by equation (14). We thus have the result that *an assembly of a large number of similar systems is described by the same equations, whose solutions are to be interpreted in the same way, whether the systems are independent or satisfy the Einstein-Bose statistics.*

Since an assembly of independent systems and an assembly satisfying the Einstein-Bose statistics are two physically different things, it may seem strange that they are both to be described by the same set of equations, even though we are restricting ourselves to the limiting case of a large number of systems in the assembly. The solution of the paradox lies in the fact that there remains an essential difference between the mathematical treatments of the two assemblies, in spite of the similarities pointed out above, as may be seen from the following discussion. An assembly of independent systems is described as completely as quantum mechanics allows when we are given the number of systems in each state. The modulus of the wave function



$(q^{(a)})$  is then determined for each state  $q^{(a)}$ , but not its phase. *This phase has no physical meaning.* We must average over all values of this phase if it appears in the result of any calculation. On the other hand, for an assembly satisfying the Einstein-Bose statistics, the  $\xi_a$ 's are dynamical variables and their phases as well as their moduli are *observables*.

There are two generalizations which we can easily make in the form of Hamiltonian which we had at the end of the preceding section for our Einstein-Bose assembly. Firstly, we may suppose that the various systems of the assembly are perturbed, not by an external field of force, but by interaction with some other atomic system, which we shall call the perturber. This will make a difference because the reaction of the assembly on the perturber will be taken into account. We must now introduce some more dynamical variables,  $\beta$  say, to describe the perturber. Our Hamiltonian will be of the form

$$H = H_P + \sum_r U_r, \quad (22)$$

where  $H_P$  is the Hamiltonian that describes the perturber alone and  $U_r$  is the energy associated with the  $r$ -th system of the assembly, consisting of its proper energy plus its interaction energy with the perturber.  $H_P$  will be a function of the  $\beta$ 's only and  $U_r$  will be a function of the variables describing the  $r$ th system and also of the  $\beta$ 's. We can express the new sum  $\sum_r U_r$  in terms of the  $n_a, w_a$  variables by the same method as before and the result will be of the same form (16), with the difference that the  $U_{ab}$ 's will no longer be numbers but will be functions of the  $\beta$ 's. The definition of  $U_{ab}$  will now be that its representative in the  $\zeta$ -representation, the  $\zeta$ 's being any complete set of commuting observables taken out of the  $\beta$ 's, is

$$(\zeta' | U_{ab} | \zeta'') = (\zeta' q_r^{(a)} | U_r | \zeta'' q_r^{(b)}), \quad (23)$$

the matrix on the right being the representative of  $U_r$  in the representation in which  $q_r$  and  $\zeta$  are diagonal. We shall still have  $U_{ab}$  commuting with the  $n$ 's and  $e^{iw}$ 's and  $e^{-iw}$ 's. The total Hamiltonian (22) will now be

$$H = H_P + \sum_{ab} \bar{\xi}_a U_{ab} \xi_b. \quad (24)$$

The second generalization which we can make is to allow the total number of systems in the assembly to vary. This generalization is necessary when the theory is applied to photons, since any emission or absorption of a photon by an atomic system results in a change in



the total number of photons in existence. We can get the theory for a varying number of systems from the theory for a fixed number by postulating in the latter theory a *zero state* for the systems, in which they are not physically observable in any way. Variations in the total number of observable systems can then be interpreted as arising from systems making transitions into or out of the zero state. We must suppose the number of systems in the zero state to be infinitely great, in order to allow the number of observable systems to increase without limit.

Using the suffix 0 to denote the zero state, we can write the Hamiltonian (24) as

$$H = H_P + \bar{\xi}_0 U_{00} \xi_0 + \sum_a \bar{\xi}_a U_{a0} \xi_0 + \sum_b \bar{\xi}_0 U_{0b} \xi_b + \sum_{ab} \bar{\xi}_a U_{ab} \xi_b, \quad (25)$$

the value 0 for  $a$  or  $b$  being excluded from the summations in (25). We may assume  $U_{00}$  to be zero, since it has no physical meaning. Since  $n_0$  is infinitely great,  $\xi_0$  and  $\bar{\xi}_0$  will also be infinitely great, of the order of  $n_0^{\frac{1}{2}}$ . The terms involving  $\xi_0$  and  $\bar{\xi}_0$  in the Hamiltonian (25), namely  $\sum_a \bar{\xi}_a U_{a0} \xi_0$  and  $\sum_b \bar{\xi}_0 U_{0b} \xi_b$ , are the terms which give rise to transitions into or out of the zero state and must be finite in order to lead to finite transition probabilities. This requires that  $U_{a0}$  and  $U_{0b}$  shall be infinitely small in such a way that  $U_{a0} \xi_0$  and  $\bar{\xi}_0 U_{0b}$  are finite. Put

$$U_{a0} \xi_0 = V_a, \quad \bar{\xi}_0 U_{0b} = \bar{V}_b, \quad (26)$$

so that  $V_a$  and  $\bar{V}_b$  are finite. The Hamiltonian (25) now becomes

$$H = H_P + \sum_a \bar{\xi}_a V_a + \sum_b \bar{V}_b \xi_b + \sum_{ab} \bar{\xi}_a U_{ab} \xi_b. \quad (27)$$

We may suppose the  $V_a$  and  $\bar{V}_b$  here to be, like the  $U_{ab}$ , functions only of the dynamical variables  $\beta$  describing the perturber and neglect their dependence on  $\xi_0$  and  $\bar{\xi}_0$  given by (26), since the P.B. of  $\xi_0$  and  $\bar{\xi}_0$  is infinitely small compared with  $\xi_0$  and  $\bar{\xi}_0$  themselves, so that  $\xi_0$  and  $\bar{\xi}_0$  may without error be counted as numbers.

#### 64. Application to Photons

In Chapter IX a theory was given of the scattering, absorption and emission of a particle by an atomic system. The interaction of the particle and atomic system was assumed to be describable by an interaction energy  $V$  appearing in the Hamiltonian, which interaction energy had to be small but was otherwise arbitrary. If we could determine the energy of interaction between a photon and an atom



or molecule, we could apply the methods of Chapter IX immediately to the case when the incident particle is a photon. We should then have a theory of the interaction of light with an atomic system. We cannot, however, determine this energy of interaction directly from analogy with the classical theory, in the way we obtained the Hamiltonians for most of the systems dealt with up to the present, since the phenomenon of the interaction of a photon with an atom has no analogue in the classical theory. We must proceed in a more indirect way. We know that the interaction of an atom with a field of radiation can be described approximately by classical electrodynamics when the field of radiation consists of a large number of photons. Our method is therefore to assume an arbitrary interaction energy  $V$  between a single photon and the atom and then in terms of  $V$  to investigate the interaction of a large number of photons with the atom. By comparing this interaction with that given by classical electrodynamics we can then obtain  $V$ .

For investigating the interaction of a large number of photons with an atom we can use the foregoing theory of an Einstein-Bose assembly, taking the photons to be the systems of the assembly and the atom to be the perturber. Since there is no interaction between photons, the total energy will be of the form (22),  $H_P$  being the energy of the atom alone and  $U_r$  the energy associated with the  $r$ -th photon, consisting of its proper energy  $h\nu_r$  together with its energy of interaction,  $V_r$  say, with the atom. Thus

$$U_r = h\nu_r + V_r. \quad (28)$$

It is convenient to take the variables  $q$  to be constants of the motion for an unperturbed photon, so that the  $q^{(a)}$ 's label the stationary states of the photons and the  $n_a$ 's are the numbers of photons in the stationary states. This requires that the  $q$ 's for a photon shall specify its momentum and polarization. Let us introduce a vector  $\mathbf{k}$ , equal to  $\hbar^{-1}$  times the momentum of a photon, and suppose the  $q$ 's for a photon to consist of its  $\mathbf{k}$  together with a polarization variable. Then each value for the suffix  $a$  in  $n_a$  and  $\xi_a$ , denoting a stationary state of a photon, will correspond to a value for  $\mathbf{k}$ , which we call  $\mathbf{k}_a$ , and a value for the polarization variable.

For each value of  $\mathbf{k}$  there are two independent stationary states, corresponding to the two independent states of polarization of a photon. We shall take these two independent stationary states to



correspond to two perpendicular states of linear polarization. We ought now to verify that the theory remains invariant under a rotation of our standard directions of polarization. Calling the two original  $\xi$ 's for the two states for the given value of  $\mathbf{k}$ ,  $\xi_1$  and  $\xi_2$ , we obtain on rotating our standard directions of polarization through an angle  $\theta$ , two new  $\xi$ 's,  $\xi_1^*$  and  $\xi_2^*$  say, given by

$$\begin{aligned}\xi_1^* &= \xi_1 \cos \theta + \xi_2 \sin \theta \\ \xi_2^* &= -\xi_1 \sin \theta + \xi_2 \cos \theta,\end{aligned}$$

since, as shown by equations (20), the transformation law for the  $\xi$ 's is the same as that for the states of polarization of a single photon. It can now easily be verified that if  $\xi_1$  and  $\xi_2$  satisfy the quantum conditions (15), then so do  $\xi_1^*$  and  $\xi_2^*$ . This is all that is necessary to establish the required invariance. We could alternatively work with the circular directions of polarization, which would mean using two  $\xi$ 's whose expressions in terms of the above ones are  $2^{-\frac{1}{2}}(\xi_1 + i\xi_2)$  and  $2^{-\frac{1}{2}}(\xi_1 - i\xi_2)$ , which again satisfy the quantum conditions (15).

Taking the representative of equation (28), we get

$$U_{ab} = h\nu_a \delta_{ab} + V_{ab},$$

$\nu_a$  being the frequency of a photon in the stationary state  $a$ . The  $V_{ab}$ 's, like the  $U_{ab}$ 's, are functions of the dynamical variables of the atom. The expression (27) for the Hamiltonian now becomes

$$H = H_P + H_R + H_Q, \quad (29)$$

where

$$H_R = \sum_a n_a h\nu_a, \quad (30)$$

the total proper energy of the radiation, and

$$\begin{aligned}H_Q &= \sum_a \{\bar{\xi}_a V_a + \bar{V}_a \xi_a\} + \sum_{ab} \bar{\xi}_a V_{ab} \xi_b \\ &= \sum_a \{V_a n_a^{\frac{1}{2}} e^{i\nu_a} + \bar{V}_a (n_a + 1)^{\frac{1}{2}} e^{-i\nu_a}\} + \sum_{ab} V_{ab} n_a^{\frac{1}{2}} e^{i\nu_a} (n_b + 1)^{\frac{1}{2}} e^{-i\nu_b}, \quad (31)\end{aligned}$$

the total interaction energy.

A photon has a continuous range of stationary states and not a discrete set, since the components of  $\mathbf{k}$  may have any values from  $-\infty$  to  $\infty$ . We therefore ought to change the sums in (31) into integrals. To do this accurately would not be very easy, since it would mean dealing according to quantum mechanics with a dynamical system with continuously many degrees of freedom, which would require a new scheme of notation and a new mathematical technique. We are, however, interested in the interaction energy (31) mainly



with regard to the limiting case of large  $n$ 's, when classical mechanics may be assumed to apply for the radiation, since we wish to compare the interaction energy in this case with that provided by classical electromagnetic theory and thus obtain expressions for the  $V_a$ 's and  $V_{ab}$ 's. In this limiting case the passage from sums to integrals is quite easy.

Let  $s_a$  denote the number of states of the photon (with a particular polarization) per unit of  $\mathbf{k}$ -space about the value  $\mathbf{k}_a$ . We assume  $s_a$  to be large, but an arbitrary function of  $\mathbf{k}_a$ , and investigate the limit of (31) when  $s_a$  is made infinite. The number of photons (with a particular polarization) per unit of  $\mathbf{k}$ -space about the value  $\mathbf{k}_a$  is

$$\eta_a = n_a s_a, \quad (32)$$

provided  $n_a$  varies in some roughly continuous way from one state to the next. Let  $(a|V|b)$  be the matrix† representing the interaction energy  $V$  for one photon in the representation for one photon, when we use the normalization rule (23) of Chapter IV for the parameter  $\mathbf{k}$ . This representation differs from the one we have used up to the present in this chapter, in which  $V$  is represented by  $V_{ab}$ , through the factor  $s$  in the weight function, according to the work at the end of § 24, so that the matrix elements in the two representations are connected by

$$(a|V|b) = V_{ab}(s_a s_b)^{\frac{1}{2}}. \quad (33)$$

Similarly, the matrix elements  $(a|V|0)$ ,  $(0|V|a)$ , referring to transitions into or out of the zero state, are connected with  $V_a$  and  $\bar{V}_a$  by

$$(a|V|0) = V_a s_a^{\frac{1}{2}} \quad (0|V|a) = \bar{V}_a s_a^{\frac{1}{2}}.$$

We can now express the interaction energy (31) in the limiting case of large  $n$ 's, when the  $n$ 's may be assumed to commute with the  $e^{i\omega}$ 's,  $e^{-i\omega}$ 's, in the form

$$\begin{aligned} H_Q &= \sum_a \{ (a|V|0) \eta_a^{\frac{1}{2}} e^{i\omega_a} + (0|V|a) \eta_a^{\frac{1}{2}} e^{-i\omega_a} \} s_a^{-1} + \\ &\quad + \sum_{ab} (a|V|b) \eta_a^{\frac{1}{2}} \eta_b^{\frac{1}{2}} e^{i(\omega_a - \omega_b)} s_a^{-1} s_b^{-1} \\ &= \sum \int \{ (a|V|0) \eta_a^{\frac{1}{2}} e^{i\omega_a} + (0|V|a) \eta_a^{\frac{1}{2}} e^{-i\omega_a} \} d\mathbf{k}_a + \\ &\quad + \sum \iint (a|V|b) \eta_a^{\frac{1}{2}} \eta_b^{\frac{1}{2}} e^{i(\omega_a - \omega_b)} d\mathbf{k}_a d\mathbf{k}_b \end{aligned} \quad (34)$$

in the limit  $s \rightarrow \infty$ , the sums in (34) referring only to the polarization.

† The matrix elements of this matrix are actually functions of the dynamical variables describing the atom, like the  $V_{ab}$ 's, and not numbers, but this does not invalidate the argument. The representation is an 'incomplete' one, whose representatives are defined in terms of those of a complete one by equations like (23).



The fact that the  $s$ 's have disappeared from this result justifies our method of dealing with a continuous range of states as a limiting case of a discrete set.

### 65. Determination of the Interaction Energy between a Photon and an Atom

We shall now determine the matrix elements  $(a|V|0)$ ,  $(0|V|a)$  and  $(a|V|b)$  by comparing (34) with the classical expression for the interaction energy between an atom and a field of radiation. For simplicity we shall suppose the atom to consist of a single electron moving in an electrostatic field of force. The field of radiation may be described by a 4-vector potential. This potential is to a certain extent arbitrary and may be chosen so that its time component vanishes. The field is then completely described by a magnetic potential  $A_x, A_y, A_z$ , or  $\mathbf{A}$ . The change that the field causes in the Hamiltonian describing the atom is now, as explained at the beginning of § 43,

$$\frac{1}{2m} \left\{ \left( \mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 - \mathbf{p}^2 \right\} = \frac{e}{mc} (\mathbf{p}, \mathbf{A}) + \frac{e^2}{2mc^2} \mathbf{A}^2. \quad (35)$$

This is the classical interaction energy, which is to be compared with (34). The  $\mathbf{A}$  that occurs here ought really to be the value of the magnetic potential at the point where the electron is momentarily situated. It is, however, a good enough approximation if we take this  $\mathbf{A}$  to be the magnetic potential at some fixed point in the atom, such as the nucleus, provided we are dealing with radiation whose wave-length is large compared with the dimensions of the atom.

To make the comparison between (34) and (35), we must first resolve the field of radiation into progressive plane-polarized trains of waves. The electric and magnetic fields of one of these trains of waves, whose frequency is  $\nu_a$  and whose direction is specified by the vector  $\mathbf{k}_a$ , are of the form

$$\mathcal{E}_a \cos[2\pi\nu_a t - (\mathbf{k}_a, \mathbf{x}) + \gamma_a], \quad \mathcal{H}_a \cos[2\pi\nu_a t - (\mathbf{k}_a, \mathbf{x}) + \gamma_a],$$

the amplitudes  $\mathcal{E}_a$  and  $\mathcal{H}_a$  being vectors of equal length that are perpendicular to the direction of motion and to each other. The total electric and magnetic fields are expressible as Fourier integrals of the form

$$\left. \begin{aligned} \mathcal{E} &= \sum \int \mathcal{E}_a \cos[2\pi\nu_a t - (\mathbf{k}_a, \mathbf{x}) + \gamma_a] d\mathbf{k}_a \\ \mathcal{H} &= \sum \int \mathcal{H}_a \cos[2\pi\nu_a t - (\mathbf{k}_a, \mathbf{x}) + \gamma_a] d\mathbf{k}_a, \end{aligned} \right\} \quad (36)$$



the  $\sum$ 's here meaning sums over both states of polarization for each value of  $\mathbf{k}_a$ .

We must obtain the distribution of energy of this field over the various Fourier components. At time  $t = 0$  we have

$$\begin{aligned}\int \mathcal{E}^2 \, d\mathbf{x} &= \sum \iint (\mathcal{E}_a, \mathcal{E}_b) \, d\mathbf{k}_a \, d\mathbf{k}_b \int \cos[\gamma_a - (\mathbf{k}_a, \mathbf{x})] \cos[\gamma_b - (\mathbf{k}_b, \mathbf{x})] \, d\mathbf{x} \\ &= \sum \iint (\mathcal{E}_a, \mathcal{E}_b) \, d\mathbf{k}_a \, d\mathbf{k}_b \, 4\pi^3 \{ \cos(\gamma_a + \gamma_b) \delta(\mathbf{k}_a + \mathbf{k}_b) + \\ &\quad + \cos(\gamma_a - \gamma_b) \delta(\mathbf{k}_a - \mathbf{k}_b) \},\end{aligned}$$

with the help of (15) of Chapter IV, the  $\sum$ 's here meaning sums over both states of polarization for each value of  $\mathbf{k}_a$  and  $\mathbf{k}_b$ . Thus

$$\int \mathcal{E}^2 \, d\mathbf{x} = 4\pi^3 \sum \int (\mathcal{E}_a, \mathcal{E}_{a'}) \cos(\gamma_a + \gamma_{a'}) \, d\mathbf{k}_a + 4\pi^3 \sum \int \mathcal{E}_a^2 \, d\mathbf{k}_a,$$

where the Fourier component specified by  $a'$  is such that  $\mathbf{k}_{a'} = -\mathbf{k}_a$ . Similarly,

$$\int \mathcal{H}^2 \, d\mathbf{x} = 4\pi^3 \sum \int (\mathcal{H}_a, \mathcal{H}_{a'}) \cos(\gamma_a + \gamma_{a'}) \, d\mathbf{k}_a + 4\pi^3 \sum \int \mathcal{H}_a^2 \, d\mathbf{k}_a.$$

On account of the connexion between the vectors  $\mathcal{E}_a$  and  $\mathcal{H}_a$  we have  $\mathcal{E}_a^2 = \mathcal{H}_a^2$  and also  $(\mathcal{E}_a, \mathcal{E}_{a'}) = -(\mathcal{H}_a, \mathcal{H}_{a'})$ . Hence the total energy is

$$1/8\pi \cdot \int (\mathcal{E}^2 + \mathcal{H}^2) \, d\mathbf{x} = \pi^2 \sum \int \mathcal{E}_a^2 \, d\mathbf{k}_a \quad (37)$$

and the energy per unit of  $\mathbf{k}$ -space for a definite polarization is  $\pi^2 \mathcal{E}_a^2$ . This may be equated to  $h\nu_a \eta_a$ , the  $\eta$  having the same meaning as in the preceding section. Thus

$$\mathcal{E}_a^2 = \pi^{-2} h\nu_a \eta_a. \quad (38)$$

The vector potential  $\mathbf{A}$  may be expressed as a Fourier integral in the same way as  $\mathcal{E}$  and  $\mathcal{H}$ . We have

$$\mathbf{A} = -\sum \int \mathbf{A}_a \sin[2\pi\nu_a t - (\mathbf{k}_a, \mathbf{x}) + \gamma_a] \, d\mathbf{k}_a, \quad (39)$$

the vector  $\mathbf{A}_a$  being in the same direction as  $\mathcal{E}_a$  and having its length given by

$$\mathbf{A}_a^2 = \left( \frac{c}{2\pi\nu_a} \right)^2 \mathcal{E}_a^2 = \frac{c^2 h}{4\pi^4 \nu_a} \eta_a. \quad (40)$$

At the origin  $\mathbf{A}$  will have the value

$$\mathbf{A} = -\sum \int \mathbf{A}_a \sin[2\pi\nu_a t + \gamma_a] \, d\mathbf{k}_a = \sum \int \mathbf{A}_a \cos w_a \, d\mathbf{k}_a,$$

$w_a$  being an angle variable of the same nature as those occurring in



(34). This value for  $\mathbf{A}$  substituted in expression (35) for the interaction energy gives

$$\begin{aligned} e/mc \cdot \sum \int (\mathbf{p}, \mathbf{A}_a) \cos w_a d\mathbf{k}_a + e^2/2mc^2 \cdot \sum \iint (\mathbf{A}_a, \mathbf{A}_b) \cos w_a \cos w_b d\mathbf{k}_a d\mathbf{k}_b \\ = \frac{eh^{\frac{1}{2}}}{2\pi^2 m} \int \frac{1}{v_a^{\frac{1}{2}}} p_a \eta_a^{\frac{1}{2}} \cos w_a d\mathbf{k}_a + \\ + \frac{e^2 h}{8\pi^4 m} \iint \frac{1}{v_a^{\frac{1}{2}} v_b^{\frac{1}{2}}} \cos \theta_{ab} \eta_a^{\frac{1}{2}} \eta_b^{\frac{1}{2}} \cos w_a \cos w_b d\mathbf{k}_a d\mathbf{k}_b, \quad (41) \end{aligned}$$

with the help of (40), where  $p_a$  is the component of the momentum  $\mathbf{p}$  of the electron in the direction of  $\mathbf{A}_a$  or  $\mathbf{e}_a$  and  $\theta_{ab}$  is the angle between the vectors  $\mathbf{A}_a$  and  $\mathbf{A}_b$ .

If we write (41) in terms of  $e^{iw}$  and  $e^{-iw}$  instead of  $\cos w$  and compare it with (34), we obtain

$$\left. \begin{aligned} (a|V|0) &= (0|V|a) = \frac{eh^{\frac{1}{2}}}{4\pi^2 m v_a^{\frac{1}{2}}} p_a \\ (a|V|b) &= \frac{e^2 h}{16\pi^4 m v_a^{\frac{1}{2}} v_b^{\frac{1}{2}}} \cos \theta_{ab}. \end{aligned} \right\} \quad (42)$$

We also find that there are certain terms in (41), namely those involving  $e^{i(w_a+w_b)}$  or  $e^{-i(w_a+w_b)}$ , which have no corresponding terms in (34). This discrepancy shows an inadequacy of the assumption that the Hamiltonian describing the interaction of an assembly of photons with an atom is of the form (22). The extra terms in (41) would give rise to transitions in which two photons are simultaneously absorbed or emitted and the possibility of such transitions requires a more complicated interaction energy than that assumed in (22). The physical effects of these terms are, however, small and unimportant, and so we shall neglect them.

Equations (42) now give the interaction energy  $V$  between a single photon and the atom. This interaction energy cannot conveniently be expressed explicitly in terms of dynamical variables. In using (42) we may, without serious error, take for the momentum  $\mathbf{p}$  of the electron its value when the atom is not perturbed by any radiation, namely  $m\dot{\mathbf{x}}$ . The left-hand sides of (42) are not 'complete' matrix elements, being functions of the dynamical variables of the atom, but we can obtain the 'complete' matrix elements from them by using formula (23). If the different stationary states of the atom alone are denoted by



$\alpha', \alpha'', \dots$ , we shall have

$$(a\alpha'|V|0\alpha'') = (0\alpha'|V|a\alpha'') = \frac{e\hbar^{\frac{1}{2}}}{4\pi^2\nu_a^{\frac{1}{2}}} (\alpha'|\dot{x}_a|\alpha'') \quad (43)$$

$$(a\alpha'|V|b\alpha'') = \frac{e^2\hbar}{16\pi^4 m\nu_a^{\frac{1}{2}}\nu_b^{\frac{1}{2}}} \cos\theta_{ab} \delta_{\alpha'\alpha''}. \quad (44)$$

Each  $a$  or  $b$  here specifies a value for  $\mathbf{k}$ , determining a momentum for the photon, and also a polarization variable determining a direction of electric force. The matrix element  $(\alpha'|\dot{x}_a|\alpha'')$  is the component of the vector  $(\alpha'|\dot{\mathbf{x}}|\alpha'')$  in the direction of the electric force specified by  $a$  and similarly  $\theta_{ab}$  is the angle between the directions of electric force specified by  $a$  and  $b$ .

## 66. Emission, Absorption and Scattering of Radiation

We can now determine directly the coefficients of emission, absorption and scattering of radiation by substituting in the formulas of Chapter IX the values for the matrix elements given by (43) and (44). These matrix elements must first be corrected by the insertion of a factor  $\hbar^{-\frac{1}{2}}$  in (43) and  $\hbar^{-3}$  in (44), owing to the different weight functions of the representation used in Chapter IX with the momentum of the incident particle labelling the representatives and the representation of § 65 with  $\mathbf{k}$ , equal to  $\hbar^{-1}$  times this momentum, labelling the representatives.

For determining the emission probability we can use formula (56) of Chapter IX. This shows that for an atom in a state  $\alpha'$  the probability per unit time per unit solid angle of its spontaneously emitting a photon and dropping to a state  $\alpha''$  of lower energy is

$$\frac{4\pi^2}{\hbar} \frac{WP}{c^2} \left| \frac{e}{\hbar} \frac{1}{(2\pi\nu)^{\frac{1}{2}}} (\alpha'|\dot{x}_a|\alpha'') \right|^2. \quad (45)$$

Now the energy and momentum of a photon of frequency  $\nu$  are

$$W = h\nu \quad P = h\nu/c.$$

Again, from the Heisenberg law (15) of Chapter VI,

$$(\alpha'|\dot{x}_a|\alpha'') = 2\pi i\nu(\alpha'\alpha'')(\alpha'|x_a|\alpha''),$$

$\nu(\alpha'\alpha'')$  being the frequency connected with transitions from state  $\alpha'$  to state  $\alpha''$ , which in the present case is just the frequency  $\nu$  of the emitted radiation. These results substituted in (45) make the emission coefficient reduce to

$$\frac{(2\pi\nu)^3}{\hbar c^3} |(\alpha'|ex_a|\alpha'')|^2. \quad (46)$$



To obtain the rate of emission of energy per unit solid angle for a specified polarization, we must multiply this by  $h\nu$ . This gives for the total rate of emission of energy in all directions

$$\frac{4}{3} \frac{(2\pi\nu)^4}{c^3} |(\alpha' | e\mathbf{X} | \alpha'')|^2, \quad (47)$$

which is in agreement with expression (27) of Chapter VIII and justifies Heisenberg's assumption for the interpretation of his matrix elements.

In the same way the absorption coefficient, given by formula (59) of Chapter IX, becomes for photons

$$\frac{4\pi^2 h^2 W}{c^2 P} \left| \frac{e}{h} \frac{1}{(2\pi\nu)^{\frac{1}{2}}} (\alpha' | \dot{x}_a | \alpha'') \right|^2 = \frac{8\pi^3 \nu}{c} |(\alpha' | ex_a | \alpha'')|^2.$$

This absorption coefficient refers to an incident beam of one photon crossing unit area per unit time per unit energy range. If we take one per unit frequency range instead of energy range, as is usual when dealing with radiation, the absorption coefficient becomes

$$\frac{8\pi^3 \nu}{hc} |(\alpha' | ex_a | \alpha'')|^2.$$

This result is the same as (25) of § 48, if we substitute for the  $E$ , there the energy  $h\nu$  of a single photon. Thus *the elementary theory of § 48, in which the radiation field is treated as an external perturbation, gives the correct value for the absorption coefficient.* The average absorption for all directions of motion and of polarization of the incident beam is

$$\frac{8\pi^3}{3} \frac{\nu}{ch} |(\alpha' | e\mathbf{X} | \alpha'')|^2,$$

which is just equal to the emission coefficient (47) divided by the factor  $8\pi h\nu^3/c^2$ . This ratio for the absorption and emission coefficients is in agreement with Einstein's theory, discussed in § 48.

Let us now consider scattering. The true scattering coefficient is given by formula (38) of Chapter IX. Such scattering of photons will not be accompanied by any change of state of the atom on account of the factor  $\delta_{\alpha'\alpha''}$  in the expression for the matrix element  $(a\alpha' | V | b\alpha'')$  in (44). Thus the final energy  $W'$  of the photon will equal its initial energy  $W^0$ . The scattering coefficient now reduces to

$$e^4/m^2 c^4 \cdot \cos^2 \theta_{ab}.$$

This is the same as that given by classical mechanics for the scattering



of radiation by a free electron. We thus see that the true scattering of radiation by an electron in an atom is independent of the atom and is correctly given by the classical theory. This result, it should be remembered, holds only provided the wave-length of the radiation is large compared with the dimensions of the atom.

The true scattering is a mathematical concept and cannot be separated out experimentally from the total scattering, given by formula (44) of Chapter IX. Let us see what this total scattering is in the case of photons. A modification must now be made in the application of formula (44) of Chapter IX. The summation  $\sum_k$  in this formula may be considered as representing the contribution to the scattering of double transitions consisting of transitions firstly from the initial state to state  $k$  and secondly from state  $k$  to the final state. The first transition may be an absorption of the incident photon and the second an emission of the required scattered photon, but it is also possible for the first transition to be the emission and the second the absorption. It is clear from the general nature of the method used for deriving formula (44) of Chapter IX that both these kinds of double transitions must be included in the summation  $\sum_k$  when this formula is applied to photons, although only the first of them was taken into account in the actual derivation given in Chapter IX.

For the double transition of absorption followed by emission we must take, using zero, single prime and double prime to refer to the initial, final and intermediate  $k$  states of the atom, respectively, and  $a$  and  $b$  to refer to the absorbed and emitted photons respectively,

$$(k|V|a\alpha^0) = (0\alpha''|V|a\alpha^0) \quad (b\alpha'|V|k) = (b\alpha'|V|0\alpha'')$$

$$E - E_k = h\nu^0 + H_P(\alpha^0) - H_P(\alpha'') = h[\nu^0 - \nu(\alpha''\alpha^0)],$$

where  $\nu^0$  is the frequency of the incident photon and

$$h\nu(\alpha''\alpha^0) = H_P(\alpha'') - H_P(\alpha^0).$$

Similarly, for the double transition of emission followed by absorption we must take

$$(k|V|a\alpha^0) = (b\alpha''|V|0\alpha^0) \quad (b\alpha'|V|k) = (0\alpha'|V|a\alpha'')$$

$$E - E_k = h\nu^0 + H_P(\alpha^0) - H_P(\alpha'') - h\nu^0 - h\nu' = -h[\nu' + \nu(\alpha''\alpha^0)],$$

where  $\nu'$  is the frequency of the scattered photon, there being now



two photons, of frequencies  $\nu^0$  and  $\nu'$ , in existence for the intermediate state  $k$ . The expression for the scattering coefficient now reduces to

$$\frac{e^4}{h^2 c^4} \frac{\nu'}{\nu^0} \left| \frac{\hbar}{m} \cos \theta_{ab} \delta_{\alpha^0 \alpha'} + \sum_{\alpha''} \left\{ \frac{(\alpha' | \dot{x}_b | \alpha'')(\alpha'' | \dot{x}_a | \alpha^0)}{\nu^0 - \nu(\alpha'' \alpha^0)} - \frac{(\alpha' | \dot{x}_a | \alpha'')(\alpha'' | \dot{x}_b | \alpha^0)}{\nu' + \nu(\alpha'' \alpha^0)} \right\} \right|^2. \quad (48)$$

If we write (48) in terms of  $x$  instead of  $\dot{x}$ , we get

$$\frac{(2\pi e)^4}{h^2 c^4} \frac{\nu'}{\nu^0} \left| \frac{\hbar}{2\pi m} \cos \theta_{ab} \delta_{\alpha^0 \alpha'} - \sum_{\alpha''} \nu(\alpha' \alpha'') \nu(\alpha'' \alpha^0) \left\{ \frac{(\alpha' | x_b | \alpha'')(\alpha'' | x_a | \alpha^0)}{\nu^0 - \nu(\alpha'' \alpha^0)} - \frac{(\alpha' | x_a | \alpha'')(\alpha'' | x_b | \alpha^0)}{\nu' + \nu(\alpha'' \alpha^0)} \right\} \right|^2. \quad (49)$$

We can simplify (49) with the help of the quantum conditions. We have

$$x_b x_a - x_a x_b = 0,$$

which gives

$$\sum_{\alpha''} \{(\alpha' | x_b | \alpha'')(\alpha'' | x_a | \alpha^0) - (\alpha' | x_a | \alpha'')(\alpha'' | x_b | \alpha^0)\} = 0, \quad (50)$$

and also

$$x_b \dot{x}_a - \dot{x}_a x_b = 1/m \cdot (x_b p_a - p_a x_b) = i\hbar/m \cdot \cos \theta_{ab},$$

which gives

$$\begin{aligned} \sum_{\alpha''} \{(\alpha' | x_b | \alpha'') \cdot \nu(\alpha'' \alpha^0)(\alpha'' | x_a | \alpha^0) - \nu(\alpha' \alpha'')(\alpha' | x_a | \alpha'') \cdot (\alpha'' | x_b | \alpha^0)\} \\ = \frac{1}{2\pi i} \cdot \frac{i\hbar}{m} \cos \theta_{ab} \delta_{\alpha^0 \alpha'} = \frac{\hbar}{2\pi m} \cos \theta_{ab} \delta_{\alpha^0 \alpha'}. \end{aligned} \quad (51)$$

Multiplying (50) by  $\nu'$  and adding to (51), we obtain

$$\begin{aligned} \sum_{\alpha''} \{(\alpha' | x_b | \alpha'')(\alpha'' | x_a | \alpha^0)[\nu' + \nu(\alpha'' \alpha^0)] - (\alpha' | x_a | \alpha'')(\alpha'' | x_b | \alpha^0)[\nu' + \nu(\alpha' \alpha'')]\} \\ = \hbar/2\pi m \cdot \cos \theta_{ab} \delta_{\alpha^0 \alpha'}. \end{aligned}$$

If we substitute this expression for  $\hbar/2\pi m \cdot \cos \theta_{ab} \delta_{\alpha^0 \alpha'}$  in (49), we obtain, after a straightforward reduction making use of identical relations between the  $\nu$ 's,

$$\frac{(2\pi e)^4}{h^2 c^4} \nu^0 \nu'^3 \left| \sum_{\alpha''} \left\{ \frac{(\alpha' | x_b | \alpha'')(\alpha'' | x_a | \alpha^0)}{\nu^0 - \nu(\alpha'' \alpha^0)} - \frac{(\alpha' | x_a | \alpha'')(\alpha'' | x_b | \alpha^0)}{\nu' + \nu(\alpha'' \alpha^0)} \right\} \right|^2. \quad (52)$$

This gives the scattering coefficient in the form of the effective area that a photon has to hit per unit solid angle of scattering. It is known as the *Kramers-Heisenberg dispersion formula*, having been first



obtained by these authors from analogies with the classical theory of dispersion.

The fact that the various terms in (49) can be combined to give the result (52) justifies the assumption made in deriving formula (44) of Chapter IX, that the matrix elements  $(a\alpha'|V|b\alpha'')$  of the interaction energy are of the second order of smallness compared with the  $(a\alpha'|V|k)$  ones, at any rate when the scattered particles are photons.

## 67. Einstein's Laws of Radiation

In the preceding section we determined the probability coefficients for absorption, emission and scattering of a photon by an atom. We were there concerned with only a single photon interacting with the atom (or at most with two), the interaction energy being given by (43) and (44). To complete our theory of radiation we require to know the laws governing the interaction of a number of photons with the atom. If the atom is exposed to an incident beam of radiation containing many photons, how do the absorption, emission and scattering probabilities depend on the intensity of the beam?

This question cannot, of course, be answered simply from a consideration of the interaction energy, defined by (43) and (44), for a single photon. We have to fall back on the general interaction energy (31) for a number of photons. We shall make use of the general result (31) of § 49, according to which a transition probability is proportional to the square of the modulus of the matrix element of the perturbing energy that refers to this transition.

Let us consider an absorption process in which the number of photons in state  $a$  is reduced from  $n_a$  to  $n_a-1$ , the atom simultaneously jumping from state  $\alpha^0$  to state  $\alpha'$ . The probability of such a process will be proportional to the square of the modulus of the matrix element

$$(n_1 n_2 \dots n_a \dots \alpha^0 | H_Q | n_1 n_2 \dots n_a - 1 \dots \alpha')$$

of the total interaction energy  $H_Q$ . The only term in the expression (31) for  $H_Q$  which can contribute to this matrix element is  $V_a n_a^{\frac{1}{2}} e^{i\nu_a}$ . This matrix element is thus proportional to  $n_a^{\frac{1}{2}}$  and the transition probability is proportional to  $n_a$ , the number of photons in the state from which the absorption takes place. It follows that the probability of an absorption process is proportional to the intensity of the incident radiation.



Similarly, for an emission process, in which the number of photons in state  $a$  is increased from  $n_a$  to  $n_a+1$ , we must consider the matrix element

$$(n_1 n_2 \dots n_a \dots \alpha^0 | H_Q | n_1 n_2 \dots n_a+1 \dots \alpha').$$

The only term in expression (31) that contributes to this is  $\bar{V}_a(n_a+1)^{\frac{1}{2}}e^{-i\omega_a}$ . This matrix element is thus proportional to  $(n_a+1)^{\frac{1}{2}}$  and the transition probability to  $n_a+1$ . In the same way a scattering process, in which the number of photons in state  $a$  is decreased from  $n_a$  to  $n_a-1$  and the number in state  $b$  is increased from  $n_b$  to  $n_b+1$ , is due to the term  $V_{ab}n_a^{\frac{1}{2}}e^{i\omega_a}(n_b+1)^{\frac{1}{2}}e^{-i\omega_b}$ , if it is a true scattering process, and to the product of the two terms  $V_a n_a^{\frac{1}{2}}e^{i\omega_a}$  and  $\bar{V}_b(n_b+1)^{\frac{1}{2}}e^{-i\omega_b}$ , if otherwise. The scattering probability is thus in any case proportional to  $n_a(n_b+1)$ . To interpret these results, we can regard a proportionality to an  $n$  as a proportionality to the intensity of the corresponding beam of radiation, but a proportionality to an  $(n+1)$  can be understood only from a study of the connexion between the discrete photon states which we are here using and the actual continuous range of states which these discrete states replace.

The work at the end of § 37 shows that a discrete state must be counted as a volume  $h^3$  of phase space for the photon. Thus a number  $n_a$  of photons in a discrete state is to be counted as a distribution of  $h^{-3}n_a$  photons per unit volume per unit of momentum space, or  $c^{-3}\nu_a^2 n_a$  per unit volume per unit solid angle per unit frequency range. This corresponds to an energy density of  $hc^{-3}\nu_a^3 n_a$  per unit solid angle per unit frequency range, or to an intensity

$$I_a = h\nu_a^3/c^2 \cdot n_a$$

per unit frequency range.

The probability for an emission process, which we found was proportional to  $n_a+1$ , is thus proportional to  $I_a+h\nu_a^3/c^2$ . This means that with no incident radiation there is still a certain amount of emission (which is, in fact, given by expression (46)), but that the emission is increased or stimulated by incident radiation in the same direction and having the same frequency (and state of polarization) as the emitted radiation under consideration. Our present theory of radiation thus completes the imperfect one of § 48, and gives a ratio for the stimulated and spontaneous emissions which is in agreement with Einstein's theory of thermodynamic equilibrium mentioned in § 48.

The probability for a scattering process from state  $a$  to state  $b$ , which we found was proportional to  $n_a(n_b+1)$ , is in the same way



proportional to  $I_a(I_b + h\nu_b^3/c^2)$ . Thus the scattering of radiation is also stimulated by incident radiation in the same direction and having the same frequency as the scattered radiation. The stimulation phenomenon is, in fact, a general one, as has been shown by Einstein and Ehrenfest† from general statistical arguments.

† Einstein and Ehrenfest, *Z. f. Physik*, **19** (1923), 301. See also Pauli, *Z. f. Physik*, **18** (1923), 272.



## RELATIVISTIC THEORY OF THE ELECTRON

**68. Relativistic Treatment of a Particle**

THE theory we have been building up and applying from Chapter II onwards is essentially a non-relativistic one. We have been working all the time with one particular Lorentz frame of reference and have not made it a requirement of the theory that its results should be independent of this frame of reference. The theory was established as an analogue of the classical non-relativistic dynamics. Let us now try to make it relativistic.

In the first place we note that the general principle of superposition of states, as given in Chapter I, is a relativistic principle. It applies to 'states' with the relativistic space-time meaning. Beyond this, though, the theory does not lend itself very well to relativistic treatment, owing to the fundamental notion of an 'observable' not fitting in very well with the requirements of relativity. The measurement of an observable, in the theory we have been dealing with up to the present, has always consisted in the measurement of some dynamical variable at some instant of time in some Lorentz frame of reference and there does not seem to be any way of generalizing this notion of an observable to make it cease to refer to a particular Lorentz frame. In consequence one cannot set up a general scheme of relativistic quantum mechanics like that of Chapter II for the non-relativistic theory. All one can do is to solve special problems in a Lorentz-invariant way. This should not be regarded as a defect of the quantum theory, since it is in perfect analogy with the classical theory. Relativistic classical mechanics does not involve any such general scheme as the contact transformation theory of non-relativistic classical mechanics, but consists in the solution of comparatively special problems.

One of the special problems that can be handled relativistically is that of the motion of a particle in an external field of force. Our non-relativistic quantum mechanics applied to this problem can be made to take a relativistic form merely by a slight change of notation. We use the representation in which the coordinates of the particle are diagonal, so that the representative of a state is  $(xyz|)$ , and adopt the Schrödinger picture, so that this representative varies with the



time  $t$  according to Schrödinger's wave equation. If we now insert the variable  $t$  explicitly in the wave function ( $xyz|$ ), so that it reads ( $xyzt|$ ), we can regard the wave function as a relativistic thing involving the four variables  $x, y, z, t$  on the same footing. Such relativistic wave functions form the basis of the present theory. The  $\psi$ -symbols will now be used for the symbolic writing of these relativistic wave functions and not of functions of  $x, y, z$  only.

The important differential operators that can operate on the  $\psi$ 's of the present theory are those representing the components of momentum

$$p_x = -i\hbar \frac{\partial}{\partial x}, \quad p_y = -i\hbar \frac{\partial}{\partial y}, \quad p_z = -i\hbar \frac{\partial}{\partial z}, \quad (1)$$

and a further one

$$W = i\hbar \frac{\partial}{\partial t} \quad (2)$$

representing the energy. Note the difference in the sign in (1) and (2), a difference which is required by relativity. The operators in (1) and (2) cannot be interpreted as observables with the same degree of generality as the operators of non-relativistic quantum mechanics, since when one of the former operates on a  $\psi$  representing a state<sup>†</sup> that actually occurs in nature and thus satisfying the wave equation, the resulting function will not *in general* satisfy the wave equation and will thus not represent any actual state.<sup>†</sup> An exception to this occurs when the momentum or energy is a constant of the motion, and such exceptions are the important practical cases when a measurement of momentum or energy is required.

## 69. The Wave Equation for the Electron

Let us consider first the case of the motion of an electron in the absence of an electromagnetic field, so that the problem is simply that of the free particle, which was discussed in §35. The Hamiltonian provided by classical mechanics for this system is given by equation (38) of §35, and leads to the wave equation

$$\{W/c - (m^2c^2 + p_x^2 + p_y^2 + p_z^2)^{1/2}\}\psi = 0, \quad (3)$$

where  $W$  and the  $p$ 's are to be interpreted as operators in accordance with equations (1) and (2). Equation (3), although it takes into account correctly the variation of the mass of the particle with its

<sup>†</sup> The word 'state' is here used in the relativistic space-time sense.



velocity, is yet unsatisfactory from the point of view of relativity, because it is very unsymmetrical between  $W$  and the  $p$ 's, so much so that one cannot generalize it in a relativistic way to the case when there is a field present. We must therefore look for a new wave equation for the free particle.

If we multiply the wave equation (3) on the left by the operator  $\{W/c + (m^2c^2 + p_x^2 + p_y^2 + p_z^2)^{1/2}\}$ , we obtain the equation

$$\{W^2/c^2 - m^2c^2 - p_x^2 - p_y^2 - p_z^2\}\psi = 0, \quad (4)$$

which is of a relativistically invariant form and may therefore more conveniently be taken as the basis of a relativistic theory. Equation (4) is not completely equivalent to equation (3) since, although every solution of (3) is also a solution of (4), the converse is not true. Only those solutions of (4) belonging to positive values for  $W$  are also solutions of (3).

The wave equation (4) is not of the form required by the general laws of the quantum theory on account of its being quadratic in  $W$ . In § 31 we deduced from quite general arguments that the wave equation must be linear in the operator  $\partial/\partial t$  or  $W$ , like equation (3) of that section. We therefore seek a wave equation that is linear in  $W$  and that is roughly equivalent to (4). In order that this wave equation shall transform in a simple way under a Lorentz transformation, we try to arrange that it shall be rational and linear in  $p_x, p_y$  and  $p_z$  as well as in  $W$ , and thus of the form

$$\{W/c + \alpha_x p_x + \alpha_y p_y + \alpha_z p_z + \beta\}\psi = 0, \quad (5)$$

where the  $\alpha$ 's and  $\beta$  are independent of  $W$  and the  $p$ 's. Since we are considering the case of no field, all points in space-time must be equivalent, so that the operator in the wave equation must not involve  $x, y, z$  or  $t$ . Thus the  $\alpha$ 's and  $\beta$  must also be independent of  $x, y, z$  and  $t$ . They must therefore denote some quite new dynamical variables, which may be pictured as describing some internal motion in the electron. We shall see later that they just describe the spin of the electron. The  $\alpha$ 's and  $\beta$  must, of course, commute with  $W$  and the  $p$ 's and also with  $x, y, z$  and  $t$ .

Multiplying (5) by the operator  $\{W/c - \alpha_x p_x - \alpha_y p_y - \alpha_z p_z - \beta\}$  on the left, we obtain

$$\{W^2/c^2 - \sum_{xyz} [\alpha_x^2 p_x^2 + (\alpha_x \alpha_y + \alpha_y \alpha_x) p_x p_y + (\alpha_x \beta + \beta \alpha_x) p_x] - \beta^2\}\psi = 0.$$



This is the same as (4) if the  $\alpha$ 's and  $\beta$  satisfy the relations

$$\begin{aligned}\alpha_x^2 &= 1, & \alpha_x \alpha_y + \alpha_y \alpha_x &= 0, \\ \beta^2 &= m^2 c^2, & \alpha_x \beta + \beta \alpha_x &= 0,\end{aligned}$$

together with the relations obtained from these by permuting  $x$ ,  $y$  and  $z$ . If we write

$$\beta = \alpha_m mc,$$

these relations may be summed up in the single one,

$$\alpha_\mu \alpha_\nu + \alpha_\nu \alpha_\mu = 2\delta_{\mu\nu} \quad (\mu, \nu = x, y, z, \text{ or } m). \quad (6)$$

The four  $\alpha$ 's all anticommute with one another and the square of each is unity.

Thus by giving suitable properties to the  $\alpha$ 's and  $\beta$  we can make the wave equation (5) equivalent to (4), in so far as the motion of the electron as a whole is concerned. We may now assume (5) is the correct relativistic wave equation for the motion of an electron in the absence of a field. This gives rise to one difficulty, however, owing to the fact that (5), like (4), is not exactly equivalent to (3), but allows solutions corresponding to negative as well as positive values of  $W$ . The former do not, of course, correspond to any actually observable motion of an electron. For the present we shall consider only the positive-energy solutions and shall leave the discussion of the negative-energy ones to § 75.

We can easily obtain a representation of the four  $\alpha$ 's. They have similar algebraic properties to the  $\sigma$ 's introduced in § 19, which  $\sigma$ 's can be represented by matrices with two rows and columns. So long as we keep to matrices with two rows and columns we cannot get a representation of more than three anticommuting quantities, and we have to go to four rows and columns to get a representation of the four anticommuting  $\alpha$ 's. It is convenient first to express the  $\alpha$ 's in terms of the  $\sigma$ 's and also of a second similar set of three anticommuting variables whose squares are unity,  $\rho_1, \rho_2, \rho_3$  say, that are independent of and commute with the  $\sigma$ 's. We may take, amongst other possibilities,

$$\alpha_x = \rho_1 \sigma_x, \quad \alpha_y = \rho_1 \sigma_y, \quad \alpha_z = \rho_1 \sigma_z, \quad \alpha_m = \rho_3, \quad (7)$$

and the  $\alpha$ 's will then satisfy all the relations (6), as may easily be verified. If we now take a representation with  $\rho_3$  and  $\sigma_z$  diagonal, we



shall get the following scheme of matrices:

$$\begin{aligned}\sigma_x &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} & \sigma_y &= \begin{pmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix} & \sigma_z &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \\ \rho_1 &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} & \rho_2 &= \begin{pmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix} & \rho_3 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.\end{aligned}$$

Corresponding to the four rows and columns, the wave function must have four components. We saw in § 61 that the spin of the electron requires the wave function to have two components. The fact that our present theory gives four is due to our wave equation (5) having twice as many solutions as it ought to have, half of them corresponding to states of negative energy.

With the help of (7), the wave equation (5) may be written in the vector form

$$\{W/c + \rho_1(\boldsymbol{\sigma}, \mathbf{p}) + \rho_3 mc\}\psi = 0. \quad (8)$$

To generalize this equation to the case when there is an electromagnetic field present, we follow the classical rule of replacing  $W$  and  $\mathbf{p}$  by  $W + eA_0$  and  $\mathbf{p} + e/c \cdot \mathbf{A}$ ,  $A_0$  and  $\mathbf{A}$  being the scalar and vector potentials of the field at the place where the electron is. This gives us the equation

$$\left\{ \frac{W}{c} + \frac{e}{c} A_0 + \rho_1 \left( \boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) + \rho_3 mc \right\} \psi = 0, \quad (9)$$

which is the fundamental wave equation of the relativistic theory of the electron. The conjugate imaginary equation

$$\phi \left\{ \frac{W}{c} + \frac{e}{c} A_0 + \rho_1 \left( \boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) + \rho_3 mc \right\} = 0 \quad (10)$$

must be treated on the same footing as (9). The operators  $W$  and  $\mathbf{p}$  in (10), which operate to the left, must be interpreted, according to § 27, as having the meanings in equations (1) and (2) with the signs reversed.

## 70. Invariance under a Lorentz Transformation

Before proceeding to discuss the physical consequences of the wave equation (9) or (10), we shall first verify that our theory really is invariant under a Lorentz transformation, or, stated more accurately,



that the physical results the theory leads to are independent of the Lorentz frame of reference used. This is not by any means obvious from the form of the wave equation (9). We have to verify that, if we write down the wave equation in a different Lorentz frame, the solutions of the new wave equation may be put into one-one correspondence with those of the original one in such a way that corresponding solutions may be assumed to represent the same state. For either Lorentz frame, the square of the modulus of the wave function, summed for the four components, should give the probability per unit volume of the electron being at any given place in that Lorentz frame. This probability is of the nature of an electric density (and will be called the electric density in future, for brevity), and its values, calculated in different Lorentz frames for wave functions representing the same state, should be connected like the time components in these frames of some 4-vector. Further, the 4-dimensional divergence of this 4-vector should vanish, signifying conservation of charge, or that the electron cannot appear or disappear in any volume without passing through the boundary.

For discussing Lorentz transformations it is convenient to put  $p_0$  for  $W/c$  and to make the convention that terms containing a repeated suffix are to be summed over the values 0,  $x$ ,  $y$ ,  $z$  for that suffix. This enables us to write equation (9) in the form

$$\{\alpha_\mu(p_\mu + e/c \cdot A_\mu) + \alpha_m mc\}\psi = 0, \quad (11)$$

$\alpha_0$  being equal to unity, and similarly we can write equation (10) in the form

$$\phi\{\alpha_\mu(p_\mu + e/c \cdot A_\mu) + \alpha_m mc\} = 0. \quad (12)$$

We now apply a Lorentz transformation and denote quantities referring to the new frame by a star. The components of the 4-vectors  $\mathbf{p}$  and  $\mathbf{A}$  will transform according to a linear law of the type

$$p_\mu = a_{\mu\nu} p_\nu^*, \quad A_\mu = a_{\mu\nu} A_\nu^*. \quad (13)$$

Substituting these expressions for  $p_\mu$  and  $A_\mu$  in equations (11) and (12), we obtain

$$\left. \begin{aligned} &\{\alpha_\mu a_{\mu\nu}(p_\nu^* + e/c \cdot A_\nu^*) + \alpha_m mc\}\psi = 0 \\ \text{and} \quad &\phi\{\alpha_\mu a_{\mu\nu}(p_\nu^* + e/c \cdot A_\nu^*) + \alpha_m mc\} = 0. \end{aligned} \right\} \quad (14)$$

We now try to bring these equations back to the form of the original (11) and (12) by introducing a new wave function  $\psi^*$ , whose four components are linear functions (with constant numerical coefficients) of the four components of the original  $\psi$ . This means that  $\psi^*$  is



connected with  $\psi$  by an equation of the type

$$\psi^* = \gamma\psi, \quad (15)$$

where  $\gamma$  is an operator like the  $\alpha$ 's, which can be represented as a matrix with four rows and columns. The conjugate imaginary equation to (15) is

$$\phi^* = \phi\bar{\gamma}. \quad (16)$$

Equations (14) will go over into the equations

$$\left. \begin{aligned} \bar{\gamma}\{\alpha_\nu(p_\nu^* + e/c \cdot A_\nu^*) + \alpha_m mc\}\psi^* &= 0 \\ \text{and} \quad \phi^*\{\alpha_\nu(p_\nu^* + e/c \cdot A_\nu^*) + \alpha_m mc\}\gamma &= 0 \end{aligned} \right\} (17)$$

provided we can choose  $\gamma$  such that

$$\bar{\gamma}\alpha_\nu\gamma = \alpha_\mu a_{\mu\nu}, \quad \bar{\gamma}\alpha_m\gamma = \alpha_m. \quad (18)$$

These equations (17) are of the same form as (11) and (12), as required, since one can divide out by the extra factors  $\bar{\gamma}$  and  $\gamma$ .

In order to verify that we can always choose  $\gamma$  to satisfy equations (18), let us first take the special case when the change of our frame of reference consists simply of a rotation through a hyperbolic angle  $\theta$  in the  $xt$  plane, so that the transformation equations for the components of a 4-vector are of the type

$$\left. \begin{aligned} p_0 &= p_0^* \cosh \theta + p_x^* \sinh \theta \\ p_x &= p_0^* \sinh \theta + p_x^* \cosh \theta \\ p_y &= p_y^*, \quad p_z = p_z^*. \end{aligned} \right\} (19)$$

The values of the  $a_{\mu\nu}$  may be written down at once from a comparison of these equations with (13). With these values for the  $a_{\mu\nu}$ , it is easy to see that equations (18) hold when we take

$$\gamma = e^{\frac{1}{2}\theta\alpha_x} = \bar{\gamma}. \quad (20)$$

We have, in fact,

$$\begin{aligned} \bar{\gamma}\alpha_0\gamma &= \bar{\gamma}\gamma = e^{\theta\alpha_x} \\ &= 1 + \theta\alpha_x + \theta^2\alpha_x^2/2! + \theta^3\alpha_x^3/3! + \dots \end{aligned}$$

On account of  $\alpha_x^2 = 1$ , this reduces to

$$\begin{aligned} \bar{\gamma}\alpha_0\gamma &= \{1 + \theta^2/2! + \dots\} + \alpha_x\{\theta + \theta^3/3! + \dots\} \\ &= \cosh \theta + \alpha_x \sinh \theta \\ &= \alpha_0 \cosh \theta + \alpha_x \sinh \theta. \end{aligned}$$

Again,  $\bar{\gamma}\alpha_x\gamma = \alpha_x\bar{\gamma}\gamma = \alpha_0 \sinh \theta + \alpha_x \cosh \theta$ .

Further,  $\bar{\gamma}\alpha_y\gamma = e^{\frac{1}{2}\theta\alpha_x}\alpha_y e^{\frac{1}{2}\theta\alpha_x} = e^{\frac{1}{2}\theta\alpha_x}e^{-\frac{1}{2}\theta\alpha_x}\alpha_y = \alpha_y$ ,



since  $\alpha_y$  anticommutes with  $\alpha_x$ , which results in  $\alpha_y f(\alpha_x) = f(-\alpha_x)\alpha_y$  for any function  $f(\alpha_x)$  of  $\alpha_x$ . Similarly,

$$\bar{\gamma}\alpha_z\gamma = \alpha_z, \quad \bar{\gamma}\alpha_m\gamma = \alpha_m.$$

Thus the five equations (18) hold with  $\gamma$  given by (20) when the  $a_{\mu\nu}$  are given by (19).

As a second typical change of the frame of reference, we may consider a rotation through an angle  $\theta$  in ordinary space about the  $x$ -axis. The transformation equations are now

$$\begin{aligned} p_0 &= p_0^* & p_x &= p_x^* \\ p_y &= p_y^* \cos \theta + p_z^* \sin \theta \\ p_z &= -p_y^* \sin \theta + p_z^* \cos \theta. \end{aligned}$$

With the new values for the  $a_{\mu\nu}$  we can easily verify that equations (18) hold with

$$\gamma = e^{-\frac{1}{2}\theta\alpha_y\alpha_z} \quad \bar{\gamma} = e^{-\frac{1}{2}\theta\alpha_z\alpha_y} = e^{\frac{1}{2}\theta\alpha_y\alpha_z},$$

the analysis being very similar to the preceding case.

If two changes of the frame of reference are made consecutively, we simply have to multiply the corresponding  $\gamma$ 's to get the  $\gamma$  for the resultant change. Now any change of the frame of reference may be built up from two rotations of the types we have considered, and hence there will always be a  $\gamma$  satisfying (18).

In this way we see that the solutions of the wave equations in the new frame of reference, equations (17), can be put into a natural one-one correspondence with those of the original wave equations (11) and (12), corresponding solutions being connected by (15) and (16), and we may assume that corresponding solutions represent the same state. It remains for us to verify that the electric density transforms like the time component of a 4-vector and that the divergence of this 4-vector vanishes.

We shall introduce the notation  $\phi_r.\psi_s$  to denote the sum of the product of each of the four components of  $\phi_r$  with the corresponding component of  $\psi_s$ . In the same way  $\phi\xi.\eta\psi$ , where  $\xi$  and  $\eta$  are any linear operators that can operate on the wave functions, will denote the sum of the product of each component of  $\phi\xi$  with the corresponding component of  $\eta\psi$ . Our new symbols of the type  $\phi\xi.\eta\psi$  are functions of  $x, y, z$  and  $t$ , and are quite distinct from the products  $\phi\xi\eta\psi$  of Chapter II, which are just numbers. It should be noted that

$$\phi.\alpha\psi = \phi\alpha.\psi \tag{21}$$



when  $\alpha$  is one of the  $\alpha$ 's in the wave equation, or more generally when it is any operator which means simply taking four linear functions (whose coefficients are numbers or functions of  $x, y, z$  and  $t$ ) of the four components of the wave function.

We can now express the electric density as  $\phi \cdot \psi$ , which is the same as  $\phi \cdot \alpha_0 \psi$  or  $\phi \alpha_0 \cdot \psi$  since  $\alpha_0 = 1$ . Let us see how the four quantities  $\phi \cdot \alpha_\mu \psi$ , with  $\mu = 0, x, y, z$ , transform under a Lorentz transformation. We have, from (15), (16) and (18),

$$\begin{aligned} \phi^* \cdot \alpha_\nu \psi^* &= \phi \bar{\gamma} \cdot \alpha_\nu \gamma \psi = \phi \cdot \bar{\gamma} \alpha_\nu \gamma \psi \\ &= \phi \cdot \alpha_\mu a_{\mu\nu} \psi = (\phi \cdot \alpha_\mu \psi) a_{\mu\nu}. \end{aligned}$$

Comparing this result with (13), we see that the four quantities  $\phi \cdot \alpha_\mu \psi$  transform like the covariant components of a 4-vector. The contravariant components will be

$$\phi \cdot \psi, \quad -\phi \cdot \alpha_x \psi, \quad -\phi \cdot \alpha_y \psi, \quad -\phi \cdot \alpha_z \psi. \quad (22)$$

This verifies that our electric density  $\phi \cdot \psi$  is the time component of a 4-vector and that the corresponding space components are  $-\phi \cdot \alpha_r \psi$  (with  $r = x, y, z$ ). These space components multiplied by the factor  $c$  give the electric current, or, stated more accurately, the probability of the electron crossing unit area per unit time.

The divergence of our 4-vector is

$$\sum_\mu \pm \frac{\partial}{\partial x_\mu} (\phi \cdot \alpha_\mu \psi), \quad (23)$$

where  $x_0$  denotes  $ct$  and the  $\pm$  sign means that the  $+$  sign is to be taken for  $\mu = 0$  and the  $-$  sign for  $\mu = x, y, z$  before one does the summation. To prove this divergence vanishes, multiply equation (11) by  $\phi$  and (12) by  $\psi$ , taking the sum over the four components in each case, and subtract. The result is

$$\phi \cdot \alpha_\mu p_\mu \psi - \phi \alpha_\mu p_\mu \cdot \psi = 0,$$

the other terms cancelling on account of (21). With the help of (1) and (2) this gives

$$\sum_\mu \pm \left\{ \phi \cdot \alpha_\mu \frac{\partial \psi}{\partial x_\mu} + \frac{\partial \phi}{\partial x_\mu} \alpha_\mu \cdot \psi \right\} = 0,$$

which just expresses the vanishing of (23). In this way we complete the proof that our theory gives consistent results in whichever frame of reference it is applied.



### 71. The Motion of a Free Electron

It is of interest to consider the motion of a free electron in the above theory according to the Heisenberg picture and to study the Heisenberg equations of motion. These equations of motion can be integrated exactly, as was first done by Schrödinger.†

As Hamiltonian we must take the expression which we get as equal to  $W$  when we put the operator on  $\psi$  in (8) equal to zero, i.e.

$$H = -c\rho_1(\boldsymbol{\sigma}, \mathbf{p}) - \rho_3 mc^2 = -c(\boldsymbol{\alpha}, \mathbf{p}) - \rho_3 mc^2. \quad (24)$$

We see at once that the momentum commutes with  $H$  and is thus a constant of the motion. Further, the  $x$ -component of the velocity is

$$\dot{x} = [x, H] = -c\alpha_x. \quad (25)$$

This result is rather surprising, as it means an altogether different relation between velocity and momentum from what one has in classical mechanics. It is connected, however, with the expressions (22) for the charge density and current. The  $\dot{x}$  given by (25) has as eigenvalues  $\pm c$ , corresponding to the eigenvalues  $\pm 1$  of  $\alpha_x$ . As  $\dot{y}$  and  $\dot{z}$  are similar, we can conclude that *a measurement of a component of the velocity of a free electron is certain to lead to the result  $\pm c$* . This conclusion is easily seen to hold also when there is a field present.

Since electrons are observed in practice to have velocities considerably less than that of light, it would seem that we have here a contradiction with experiment. The contradiction is not real, though, since the theoretical velocity in the above conclusion is the velocity at one instant of time while observed velocities are always average velocities through appreciable time intervals. We shall find upon further examination of the equations of motion that the velocity is not at all constant, but oscillates rapidly about a mean value which agrees with the observed value.

It may easily be verified that a measurement of a component of the velocity must lead to the result  $\pm c$  in a relativistic theory, simply from an elementary application of the principle of uncertainty of § 28. To measure the velocity we must measure the position at two slightly different times and then divide the change of position by the time interval. (It will not do to measure the momentum and apply a formula, as the ordinary connexion between velocity and momentum is not valid.) In order that our measured velocity may approximate to the instantaneous velocity, the time interval between the

† Schrödinger, *Sitzungsb. d. Berlin Akad.*, 1930, p. 418.



two measurements of position must be very short and hence these measurements must be very accurate. The great accuracy with which the position of the electron is known during the time interval must give rise, according to the principle of uncertainty, to an almost complete indeterminacy in its momentum. This means that almost all values of the momentum are equally probable, so that the momentum is almost certain to be infinite. An infinite value for a component of momentum corresponds to the value  $\pm c$  for the corresponding component of velocity.

Let us now examine how the velocity of the electron varies with time. We have

$$i\hbar\dot{\alpha}_x = \alpha_x H - H\alpha_x.$$

Now since  $\alpha_x$  anticommutes with all the terms in  $H$  except  $-c\alpha_x p_x$ ,

$$\alpha_x H + H\alpha_x = -\alpha_x c\alpha_x p_x - c\alpha_x p_x \alpha_x = -2cp_x,$$

and hence

$$\left. \begin{aligned} i\hbar\dot{\alpha}_x &= 2\alpha_x H + 2cp_x \\ &= -2H\alpha_x - 2cp_x. \end{aligned} \right\} (26)$$

Since  $H$  and  $p_x$  are constants, it follows from the first of equations (26) that

$$i\hbar\ddot{\alpha}_x = 2\dot{\alpha}_x H. \quad (27)$$

This differential equation in  $\dot{\alpha}_x$  can be integrated immediately, the result being

$$\dot{\alpha}_x = \dot{\alpha}_x^0 e^{-2iHt/\hbar}, \quad (28)$$

where  $\dot{\alpha}_x^0$  is a constant, equal to the value of  $\dot{\alpha}_x$  when  $t = 0$ . The factor  $e^{-2iHt/\hbar}$  must be put to the right of the factor  $\dot{\alpha}_x^0$  in (28) on account of the  $H$  occurring to the right of the  $\dot{\alpha}_x$  in (27). The second of equations (26) leads in the same way to the result

$$\dot{\alpha}_x = e^{2iHt/\hbar} \dot{\alpha}_x^0.$$

We can now easily complete the integration of the equation of motion for  $x$ . From (28) and the first of equations (26)

$$\alpha_x = \frac{1}{2}i\hbar\dot{\alpha}_x^0 e^{-2iHt/\hbar} H^{-1} - cp_x H^{-1}, \quad (29)$$

and hence the time-integral of equation (25) is

$$x = \frac{1}{4}c\hbar^2\dot{\alpha}_x^0 e^{-2iHt/\hbar} H^{-2} + c^2 p_x H^{-1}t + a_x, \quad (30)$$

$a_x$  being a constant.

From (29) we see that the  $x$  component of velocity,  $-c\alpha_x$ , consists of two parts, a constant part  $c^2 p_x H^{-1}$ , connected with the momentum by the classical relativistic formula, and an oscillatory part

$$-\frac{1}{2}ic\hbar\dot{\alpha}_x^0 e^{-2iHt/\hbar} H^{-1},$$

whose frequency is high, being  $2H/\hbar$ , which is at least  $2mc^2/\hbar$ . Only



the constant part would be observed in a practical measurement of velocity, such a measurement giving the average velocity through a time-interval much larger than  $\hbar/2mc^2$ . The oscillatory part secures that the instantaneous value of  $\dot{x}$  shall have the eigenvalues  $\pm c$ . The oscillatory part of  $x$  is small, being, according to (30),

$$\frac{1}{4}c\hbar^2\ddot{\alpha}_x^0 e^{-2iH\hbar/\hbar}H^{-2} = -\frac{1}{2}ic\hbar(\alpha_x + cp_x H^{-1})H^{-1},$$

which is of the order of magnitude  $\hbar/mc$ , since  $(\alpha_x + cp_x H^{-1})$  is of the order of magnitude unity.

## 72. Existence of the Spin

In § 69 we saw that the correct wave equation for the electron in the absence of an electromagnetic field, namely equation (5) or (8), is equivalent to the wave equation (4) which is suggested from analogy with the classical theory. This equivalence no longer holds when there is a field. By treating the correct wave equation for this case, namely (9), in the same way as we treated (5) and comparing it with the wave equation to be expected from analogy with the classical theory, namely

$$\left\{\left(\frac{W}{c} + \frac{e}{c}A_0\right)^2 - \left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right)^2 - m^2c^2\right\}\psi = 0, \quad (31)$$

in which the operator is just the classical relativistic Hamiltonian, we may expect to get an indication of the new physical features of the present theory.

We must multiply (9) by some factor on the left to make it resemble (31) as closely as possible. Taking this factor to be

$$\frac{W}{c} + \frac{e}{c}A_0 - \rho_1\left(\boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c}\mathbf{A}\right) - \rho_3 mc,$$

we get

$$\left\{\left(\frac{W}{c} + \frac{e}{c}A_0\right)^2 - \left(\boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c}\mathbf{A}\right)^2 - m^2c^2 + \rho_1\left[\left(\frac{W}{c} + \frac{e}{c}A_0\right)\left(\boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c}\mathbf{A}\right) - \left(\boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c}\mathbf{A}\right)\left(\frac{W}{c} + \frac{e}{c}A_0\right)\right]\right\}\psi = 0. \quad (32)$$

We now use the general formula that, if  $\mathbf{B}$  and  $\mathbf{C}$  are any two vectors that commute with  $\boldsymbol{\sigma}$ ,

$$\begin{aligned} (\boldsymbol{\sigma}, \mathbf{B})(\boldsymbol{\sigma}, \mathbf{C}) &= \sum_{xyz} \{\sigma_x^2 B_x C_x + \sigma_x \sigma_y B_x C_y + \sigma_y \sigma_x B_y C_x\} \\ &= (\mathbf{B}, \mathbf{C}) + i \sum_{xyz} \sigma_z (B_x C_y - B_y C_x) \\ &= (\mathbf{B}, \mathbf{C}) + i(\boldsymbol{\sigma}, \mathbf{B} \times \mathbf{C}). \end{aligned} \quad (33)$$



Taking  $\mathbf{B} = \mathbf{C} = \mathbf{p} + e/c \cdot \mathbf{A}$ , we find, since

$$\begin{aligned} \left(\mathbf{p} + \frac{e}{c} \mathbf{A}\right) \times \left(\mathbf{p} + \frac{e}{c} \mathbf{A}\right) &= \frac{e}{c} \{\mathbf{p} \times \mathbf{A} + \mathbf{A} \times \mathbf{p}\} \\ &= -i\hbar e/c \cdot \text{curl} \mathbf{A} = -i\hbar e/c \cdot \mathcal{H}, \end{aligned}$$

where  $\mathcal{H}$  is the magnetic field, that

$$\left(\boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c} \mathbf{A}\right)^2 = \left(\mathbf{p} + \frac{e}{c} \mathbf{A}\right)^2 + \frac{\hbar e}{c} (\boldsymbol{\sigma}, \mathcal{H}).$$

Also we have

$$\begin{aligned} \left(\frac{W}{c} + \frac{e}{c} A_0\right) \left(\boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c} \mathbf{A}\right) - \left(\boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c} \mathbf{A}\right) \left(\frac{W}{c} + \frac{e}{c} A_0\right) \\ = \frac{e}{c} \left(\boldsymbol{\sigma}, \frac{W}{c} \mathbf{A} - \mathbf{A} \frac{W}{c} + A_0 \mathbf{p} - \mathbf{p} A_0\right) \\ = \frac{i\hbar e}{c} \left(\boldsymbol{\sigma}, \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} + \text{grad} A_0\right) = -i \frac{\hbar e}{c} (\boldsymbol{\sigma}, \mathcal{E}), \end{aligned}$$

where  $\mathcal{E}$  is the electric field. Thus (32) becomes

$$\left\{ \left(\frac{W}{c} + \frac{e}{c} A_0\right)^2 - \left(\mathbf{p} + \frac{e}{c} \mathbf{A}\right)^2 - m^2 c^2 - \frac{\hbar e}{c} (\boldsymbol{\sigma}, \mathcal{H}) - i\rho_1 \frac{\hbar e}{c} (\boldsymbol{\sigma}, \mathcal{E}) \right\} \psi = 0.$$

This equation differs from (31) through having two extra terms in the operator. The electron according to the present theory is more closely analogous to a classical system with the Hamiltonian function

$$\left(\frac{W}{c} + \frac{e}{c} A_0\right)^2 - \left(\mathbf{p} + \frac{e}{c} \mathbf{A}\right)^2 - m^2 c^2 - \frac{\hbar e}{c} (\boldsymbol{\sigma}, \mathcal{H}) - i\rho_1 \frac{\hbar e}{c} (\boldsymbol{\sigma}, \mathcal{E}).$$

If we neglect relativistic corrections, so that we can put  $W = mc^2 + W_1$  and count  $W_1$  as small, this Hamiltonian reduces, after division throughout by  $2m$ , to

$$W_1 - \left\{ -eA_0 + \frac{1}{2m} \left(\mathbf{p} + \frac{e}{c} \mathbf{A}\right)^2 + \frac{\hbar e}{2mc} (\boldsymbol{\sigma}, \mathcal{H}) + i\rho_1 \frac{\hbar e}{2mc} (\boldsymbol{\sigma}, \mathcal{E}) \right\}.$$

We can now see that the two extra terms may be considered approximately as due to the electron possessing an additional potential energy of amount

$$\frac{\hbar e}{2mc} (\boldsymbol{\sigma}, \mathcal{H}) + i\rho_1 \frac{\hbar e}{2mc} (\boldsymbol{\sigma}, \mathcal{E}),$$

which may be interpreted as arising from the electron having a magnetic moment  $-\hbar e/2mc \cdot \boldsymbol{\sigma}$  and an electric moment  $-i\rho_1 \hbar e/2mc \cdot \boldsymbol{\sigma}$ . This



magnetic moment is in agreement with the assumption of § 43 and is what is required by experiment. The electric moment, on the other hand, is a pure imaginary quantity and thus cannot be considered as having a physical meaning. The Hamiltonian of our original wave equation (9) is real, and the imaginary term has appeared only on account of our having performed a rather artificial operation to get a Hamiltonian that can be compared with the classical one.

The spin angular momentum does not give rise to any potential energy and therefore does not appear in the result of the preceding calculation. The simplest way of showing the existence of the spin angular momentum is to take the case of the motion of a free electron or an electron in a central field of force and determine the angular momentum integrals. This means working with the Hamiltonian (24), or with this Hamiltonian generalized by the addition to it of a potential energy  $-eA_0$  which may be any function of the radius  $r$ , thus

$$H = -eA_0(r) - c\rho_1(\boldsymbol{\sigma}, \mathbf{p}) - \rho_3 mc^2. \quad (34)$$

With either Hamiltonian we find for the rate of change of the  $x$ -component of orbital angular momentum,  $m_x = yp_z - zp_y$ , with the help of commutability relations proved in §§ 38 and 40,

$$\begin{aligned} i\hbar \dot{m}_x &= m_x H - H m_x \\ &= -c\rho_1 \{m_x(\boldsymbol{\sigma}, \mathbf{p}) - (\boldsymbol{\sigma}, \mathbf{p})m_x\} \\ &= -c\rho_1(\boldsymbol{\sigma}, m_x \mathbf{p} - \mathbf{p}m_x) \\ &= -i\hbar c\rho_1 \{\sigma_y p_z - \sigma_z p_y\}. \end{aligned}$$

Thus  $\dot{m}_x \neq 0$  and the orbital angular momentum is not a constant of the motion. This result is to be expected from the integrated equation of motion (30), the oscillatory part of the motion here displayed giving rise to an oscillatory term in the angular momentum.

As a further equation of motion with the Hamiltonian (24) or (34), we have

$$\begin{aligned} i\hbar \dot{\sigma}_x &= \sigma_x H - H \sigma_x \\ &= -c\rho_1 \{\sigma_x(\boldsymbol{\sigma}, \mathbf{p}) - (\boldsymbol{\sigma}, \mathbf{p})\sigma_x\} \\ &= -c\rho_1(\sigma_x \boldsymbol{\sigma} - \boldsymbol{\sigma} \sigma_x, \mathbf{p}) \\ &= -2ic\rho_1 \{\sigma_z p_y - \sigma_y p_z\} \end{aligned}$$

with the help of equations (55) of § 19. Hence

$$i\hbar(\dot{m}_x + \frac{1}{2}\hbar \dot{\sigma}_x) = 0,$$

so that the vector  $\mathbf{m} + \frac{1}{2}\hbar \boldsymbol{\sigma}$  is a constant of the motion. This result one can interpret by saying *the electron has a spin angular momentum*



$\frac{1}{2}\hbar\sigma$ , which must be added to the orbital angular momentum  $\mathbf{m}$  before one gets a constant of the motion. The same vector  $\sigma$  fixes the directions of both the spin magnetic moment and the spin angular momentum. If an electron in a certain state of spin has a spin angular momentum of  $\frac{1}{2}\hbar$  in a particular direction, it will have a magnetic moment  $-e\hbar/2mc$  in the same direction.

### 73. Transition to Polar Variables

For the further study of the motion of an electron in a central field of force with the Hamiltonian (34), it is convenient to make a transformation to polar coordinates, as was done in § 40 in the non-relativistic case. We can introduce  $r$  and  $p_r$  as before, but instead of  $k$ , the magnitude of the orbital angular momentum  $\mathbf{m}$ , which is no longer a constant of the motion, we must now use the magnitude of the total angular momentum  $\mathbf{M} = \mathbf{m} + \frac{1}{2}\hbar\sigma$ . Let us put

$$j^2\hbar^2 = M_x^2 + M_y^2 + M_z^2 + \frac{1}{4}\hbar^2. \quad (35)$$

The eigenvalues of  $m_z$  are integral multiples of  $\hbar$ , those of  $\frac{1}{2}\hbar\sigma$  are  $\pm\frac{1}{2}\hbar$ , and hence those of  $M_z$  must be half-odd integral multiples of  $\hbar$ . It follows from the theory of § 39 that the eigenvalues of  $j$  must be integers.

If in formula (33) we take  $\mathbf{B} = \mathbf{C} = \mathbf{m}$ , we get

$$\begin{aligned} (\sigma, \mathbf{m})^2 &= \mathbf{m}^2 + i(\sigma, \mathbf{m} \times \mathbf{m}) \\ &= \mathbf{m}^2 - \hbar(\sigma, \mathbf{m}) \\ &= (\mathbf{m} + \tfrac{1}{2}\hbar\sigma)^2 - 2\hbar(\sigma, \mathbf{m}) - \tfrac{3}{4}\hbar^2. \end{aligned}$$

Hence  $\{(\sigma, \mathbf{m}) + \hbar\}^2 = \mathbf{M}^2 + \frac{1}{4}\hbar^2$ .

Thus  $(\sigma, \mathbf{m}) + \hbar$  is a quantity whose square is  $\mathbf{M}^2 + \frac{1}{4}\hbar^2$  and we could, consistently with equation (35), define  $j\hbar$  as  $(\sigma, \mathbf{m}) + \hbar$ . This would not be the most convenient definition for  $j$ , however, since we would like to have  $j$  a constant of the motion and  $(\sigma, \mathbf{m}) + \hbar$  is not constant. We have, in fact, from applications of (33),

$$(\sigma, \mathbf{m})(\sigma, \mathbf{p}) = i(\sigma, \mathbf{m} \times \mathbf{p})$$

and

$$(\sigma, \mathbf{p})(\sigma, \mathbf{m}) = i(\sigma, \mathbf{p} \times \mathbf{m}),$$

so that

$$\begin{aligned} (\sigma, \mathbf{m})(\sigma, \mathbf{p}) + (\sigma, \mathbf{p})(\sigma, \mathbf{m}) &= i \sum_{xyz} \sigma_x \{m_y p_z - m_z p_y + p_y m_z - p_z m_y\} \\ &= i \sum_{xyz} \sigma_x \cdot 2i\hbar p_x = -2\hbar(\sigma, \mathbf{p}), \end{aligned}$$

or  $\{(\sigma, \mathbf{m}) + \hbar\}(\sigma, \mathbf{p}) + (\sigma, \mathbf{p})\{(\sigma, \mathbf{m}) + \hbar\} = 0$ .



Thus  $(\boldsymbol{\sigma}, \mathbf{m}) + \hbar$  anticommutes with one of the terms in the expression (34) for  $H$ , namely the term  $-c\rho_1(\boldsymbol{\sigma}, \mathbf{p})$ , and commutes with the other two. It follows that  $\rho_3\{(\boldsymbol{\sigma}, \mathbf{m}) + \hbar\}$  commutes with all the three terms in  $H$  and is a constant of the motion. But the square of  $\rho_3\{(\boldsymbol{\sigma}, \mathbf{m}) + \hbar\}$  is also  $M^2 + \frac{1}{4}\hbar^2$ . We can therefore take

$$j\hbar = \rho_3\{(\boldsymbol{\sigma}, \mathbf{m}) + \hbar\}, \quad (36)$$

which gives us a convenient rational definition for  $j$  which is consistent with (35) and makes  $j$  a constant of the motion. The eigenvalues of this  $j$  are all positive and negative integers, excluding zero.

By a further application of (33), we get

$$\begin{aligned} (\boldsymbol{\sigma}, \mathbf{x})(\boldsymbol{\sigma}, \mathbf{p}) &= (\mathbf{x}, \mathbf{p}) + i(\boldsymbol{\sigma}, \mathbf{m}) \\ &= rp_r + i\rho_3 j\hbar, \end{aligned} \quad (37)$$

with the help of (36) and also of equation (13) of Chapter VII. We introduce the dynamical variable  $\epsilon$  defined by

$$r\epsilon = \rho_1(\boldsymbol{\sigma}, \mathbf{x}). \quad (38)$$

Since  $r$  commutes with  $\rho_1$  and with  $(\boldsymbol{\sigma}, \mathbf{x})$ , it must commute with  $\epsilon$ . We thus have

$$r^2\epsilon^2 = [\rho_1(\boldsymbol{\sigma}, \mathbf{x})]^2 = (\boldsymbol{\sigma}, \mathbf{x})^2 = \mathbf{x}^2 = r^2,$$

or

$$\epsilon^2 = 1.$$

Since there is symmetry between  $\mathbf{x}$  and  $\mathbf{p}$  so far as angular momentum is concerned,  $\rho_1(\boldsymbol{\sigma}, \mathbf{x})$ , like  $\rho_1(\boldsymbol{\sigma}, \mathbf{p})$ , must commute with  $M$  and  $j$ . Hence  $\epsilon$  commutes with  $M$  and  $j$ . Further,  $\epsilon$  must commute with  $p_r$ , since we have

$$(\boldsymbol{\sigma}, \mathbf{x})(\mathbf{x}, \mathbf{p}) - (\mathbf{x}, \mathbf{p})(\boldsymbol{\sigma}, \mathbf{x}) = (\boldsymbol{\sigma}, \mathbf{x}(\mathbf{x}, \mathbf{p}) - (\mathbf{x}, \mathbf{p})\mathbf{x}) = i\hbar(\boldsymbol{\sigma}, \mathbf{x}),$$

$$\text{which gives } r\epsilon(rp_r + i\hbar) - (rp_r + i\hbar)r\epsilon = i\hbar r\epsilon$$

$$\text{or } r\epsilon(p_r r + 2i\hbar) - (rp_r + i\hbar)r\epsilon = i\hbar r\epsilon,$$

$$\text{which reduces to } \epsilon p_r - p_r \epsilon = 0.$$

From (37) and (38) we obtain

$$r\epsilon\rho_1(\boldsymbol{\sigma}, \mathbf{p}) = rp_r + i\rho_3 j\hbar$$

$$\text{or } \rho_1(\boldsymbol{\sigma}, \mathbf{p}) = \epsilon p_r + i\epsilon\rho_3 j\hbar/r.$$

Thus (34) becomes

$$H/c = -e/c \cdot A_0 - \epsilon p_r - i\epsilon\rho_3 j\hbar/r - \rho_3 mc.$$

This gives our Hamiltonian expressed in terms of polar variables. It should be noticed that  $\epsilon$  and  $\rho_3$  commute with all the other variables occurring in  $H$  and anticommute with one another. This means that



we can take a representation in which  $\epsilon$  and  $\rho_3$  are represented respectively by the matrices

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (39)$$

and in which  $r$ , say, is diagonal, and the representative  $(r|)$  of a state will then have two components,  $(r|)_a$  and  $(r|)_b$  say, referring to the two rows and columns of the matrices (39).

#### 74. The Fine-Structure of the Energy-Levels of Hydrogen

We shall now take the case of the hydrogen atom, for which  $A_0 = e/r$ , and work out its energy-levels, given by the eigenvalues  $H'$  of  $H$ . The equation  $(H' - H)\psi = 0$  which defines these eigenvalues, when written in terms of representatives in the representation discussed above with  $\epsilon$  and  $\rho_3$  represented by the matrices (39), gives the equations

$$\begin{aligned} \left(\frac{H'}{c} + \frac{e^2}{cr}\right)(r|)_a - \hbar \frac{\partial}{\partial r}(r|)_b - \frac{j\hbar}{r}(r|)_b + mc(r|)_a &= 0 \\ \left(\frac{H'}{c} + \frac{e^2}{cr}\right)(r|)_b + \hbar \frac{\partial}{\partial r}(r|)_a - \frac{j\hbar}{r}(r|)_a - mc(r|)_b &= 0. \end{aligned}$$

If we put  $\frac{\hbar}{mc + H'/c} = a_1, \quad \frac{\hbar}{mc - H'/c} = a_2, \quad (40)$

these equations reduce to

$$\left. \begin{aligned} \left(\frac{1}{a_1} + \frac{\alpha}{r}\right)(r|)_a - \left(\frac{\partial}{\partial r} + \frac{j}{r}\right)(r|)_b &= 0 \\ \left(-\frac{1}{a_2} + \frac{\alpha}{r}\right)(r|)_b + \left(\frac{\partial}{\partial r} - \frac{j}{r}\right)(r|)_a &= 0, \end{aligned} \right\} \quad (41)$$

where  $\alpha = e^2/\hbar c$ , which is a small number. We shall solve these equations by a similar method to that used for equation (20) in § 41.

Put

$$(r|)_a = e^{-r/af}, \quad (r|)_b = e^{-r/ag},$$

introducing two new functions,  $f$  and  $g$ , of  $r$ , where

$$a = (a_1 a_2)^{\frac{1}{2}} = \hbar(m^2 c^2 - H'^2/c^2)^{-\frac{1}{2}}. \quad (42)$$

Equations (41) become

$$\left. \begin{aligned} \left(\frac{1}{a_1} + \frac{\alpha}{r}\right)f - \left(\frac{\partial}{\partial r} - \frac{1}{a} + \frac{j}{r}\right)g &= 0 \\ \left(-\frac{1}{a_2} + \frac{\alpha}{r}\right)g + \left(\frac{\partial}{\partial r} - \frac{1}{a} - \frac{j}{r}\right)f &= 0. \end{aligned} \right\} \quad (43)$$



We now try for a solution in which  $f$  and  $g$  are in the form of power series,

$$f = \sum_s c_s r^s, \quad g = \sum_s c'_s r^s, \quad (44)$$

in which consecutive values of  $s$  differ by unity though these values need not be integers. Substituting these expressions for  $f$  and  $g$  in (43) and picking out coefficients of  $r^{s-1}$ , we obtain

$$\left. \begin{aligned} c_{s-1}/a_1 + \alpha c_s - (s+j)c'_s + c'_{s-1}/a &= 0 \\ -c'_{s-1}/a_2 + \alpha c'_s + (s-j)c_s - c_{s-1}/a &= 0. \end{aligned} \right\} (45)$$

By multiplying the first of these equations by  $a$  and the second by  $a_2$  and adding, we can eliminate both  $c_{s-1}$  and  $c'_{s-1}$ , since from (42)  $a/a_1 = a_2/a$ . This gives

$$[a\alpha + a_2(s-j)]c_s + [a_2\alpha - a(s+j)]c'_s = 0, \quad (46)$$

a relation which shows the connexion between the primed and unprimed  $c$ 's.

The boundary condition at  $r = 0$  requires that the series (44) shall terminate on the side of small  $s$ . If  $s_0$  is the minimum value of  $s$  for which  $c_s$  and  $c'_s$  do not both vanish, we obtain from (45), by putting  $s = s_0$  and  $c_{s_0-1} = c'_{s_0-1} = 0$ ,

$$\begin{aligned} \alpha c_{s_0} - (s_0+j)c'_{s_0} &= 0 \\ \alpha c'_{s_0} + (s_0-j)c_{s_0} &= 0, \end{aligned}$$

which give  $\alpha^2 = -s_0^2 + j^2$ .

Since the boundary condition requires that the minimum value of  $s$  shall be greater than zero, we must take

$$s_0 = +\sqrt{(j^2 - \alpha^2)}.$$

To investigate the convergence of the series (44) we shall determine the ratio  $c_s/c_{s-1}$  for large  $s$ . Equation (46) and the second of equations (45) give approximately, when  $s$  is large,

$$a_2 c_s = a c'_s$$

and  $sc_s = c_{s-1}/a + c'_{s-1}/a_2$ .

Hence  $c_s/c_{s-1} = 2/as$ .

The series (44) will therefore converge like

$$\sum_s \frac{1}{s!} \left( \frac{2r}{a} \right)^s$$

or  $e^{2r/a}$ . This result is similar to that obtained in § 41 and allows us to infer, as in § 41, that all values of  $H'$  are permissible for which  $a$  is



pure imaginary, i.e., from (42), for which  $H' > mc^2$ , but of those values of  $H'$  for which  $\alpha$  is real, only those are permissible for which the series (44) terminate on the side of large  $s$ .

If the series (44) terminate with the terms  $c_s$  and  $c'_s$ , so that  $c_{s+1} = c'_{s+1} = 0$ , we obtain from (45) with  $s+1$  substituted for  $s$

$$\begin{aligned} c_s/a_1 + c'_s/a &= 0 \\ -c'_s/a_2 - c_s/a &= 0. \end{aligned}$$

These two equations are equivalent on account of (42). When combined with (46), they give

$$a_1[a\alpha + a_2(s-j)] = a[a_2\alpha - a(s+j)],$$

which reduces to  $2a_1a_2s = a(a_2 - a_1)\alpha$

or 
$$\frac{s}{a} = \frac{1}{2} \left( \frac{1}{a_1} - \frac{1}{a_2} \right) \alpha = \frac{H'}{ch} \alpha,$$

with the help of (40). Squaring and using (42), we obtain

$$s^2(m^2c^2 - H'^2/c^2) = \alpha^2 H'^2/c^2.$$

Hence

$$\frac{H'}{mc^2} = \left( 1 + \frac{\alpha^2}{s^2} \right)^{-\frac{1}{2}}.$$

The  $s$  here, which specifies the last term in the series, must be greater than  $s_0$  by some integer not less than zero. Calling this integer  $n$ , we have

$$s = n + \sqrt{(j^2 - \alpha^2)}$$

and thus 
$$\frac{H'}{mc^2} = \left\{ 1 + \frac{\alpha^2}{\{n + \sqrt{(j^2 - \alpha^2)}\}^2} \right\}^{-\frac{1}{2}}. \quad (47)$$

This formula gives the discrete energy-levels of the hydrogen spectrum and was first obtained by Sommerfeld working with Bohr's orbit theory. There are two quantum numbers  $n$  and  $j$  involved, but owing to  $\alpha^2$  being very small the energy depends almost entirely on  $n + |j|$ . Values of  $n$  and  $|j|$  that give the same  $n + |j|$  give rise to a set of energy-levels lying very close to one another, and to the energy-level given by the non-relativistic formula (27) of § 41 with  $s = n + |j|$ .

For a general value of  $n$ ,  $j$  can have any integral value except zero. The value  $n = 0$  is, however, exceptional as it makes equation (46) vanish identically. A closer investigation shows that in this case only negative values for  $j$  are allowed.†

† See W. Gordon, *Z. f. Physik*, 48 (1928), 11.



## 75. Theory of the Positron

It has been mentioned in § 69 that the wave equation for the electron admits of twice as many solutions as it ought to, half of them referring to states with negative values for the kinetic energy  $W + eA_0$ . This difficulty was introduced as soon as we passed from equation (3) to equation (4) and is inherent in any relativistic theory. It occurs also in classical relativistic theory, but is not then serious since, owing to the continuity in the variation of all classical dynamical variables, if the kinetic energy  $W + eA_0$  is initially positive (when it must be greater than or equal to  $mc^2$ ), it cannot subsequently be negative (when it would have to be less than or equal to  $-mc^2$ ). In the quantum theory, however, discontinuous transitions may take place, so that if the electron is initially in a state of positive kinetic energy it may make a transition to a state of negative kinetic energy. It is therefore no longer permissible simply to ignore the negative-energy states, as one can do in the classical theory.

Let us examine the negative-energy solutions of the equation

$$\left\{ \left( \frac{W}{c} + \frac{e}{c} A_0 \right) + \alpha_x \left( p_x + \frac{e}{c} A_x \right) + \right. \\ \left. + \alpha_y \left( p_y + \frac{e}{c} A_y \right) + \alpha_z \left( p_z + \frac{e}{c} A_z \right) + \alpha_m mc \right\} \psi = 0 \quad (48)$$

a little more closely. For this purpose it is convenient to use a representation of the  $\alpha$ 's in which all the elements of the matrices representing  $\alpha_x$ ,  $\alpha_y$  and  $\alpha_z$  are real and all those of the matrix representing  $\alpha_m$  are pure imaginary. Such a representation may be obtained, for instance, from that of § 69 by interchanging the expressions for  $\alpha_y$  and  $\alpha_m$  in (7). With such a representation, if we write  $-i$  for  $i$  in the operator of equation (48), we get, remembering (1) and (2),

$$\left\{ \left( -\frac{W}{c} + \frac{e}{c} A_0 \right) + \alpha_x \left( -p_x + \frac{e}{c} A_x \right) + \right. \\ \left. + \alpha_y \left( -p_y + \frac{e}{c} A_y \right) + \alpha_z \left( -p_z + \frac{e}{c} A_z \right) - \alpha_m mc \right\} \psi = 0. \quad (49)$$

Thus the conjugate complex of any wave function that is a solution of (48) is a solution of (49). Further, if the solution of (48) belongs to a negative value for  $W + eA_0$ , the conjugate complex solution of (49) will belong to a positive value for  $W - eA_0$ . But equation (49) is just what one would get if one substituted  $-e$  for  $e$  in (48). It follows



that the conjugate complex of any solution of (48) belonging to a negative value for  $W + eA_0$  is a solution, belonging to a positive value for  $W - eA_0$ , of the wave equation obtained from (48) by substitution of  $-e$  for  $e$ , and therefore represents an electron of charge  $+e$ , instead of  $-e$ , moving through the given electromagnetic field. Thus the unwanted solutions of (48) are connected with the motion of an electron with a charge  $+e$ . (It is not possible, of course, with an arbitrary electromagnetic field, to separate the solutions of (48) definitely into those referring to positive and those referring to negative values for  $W + eA_0$ , as such a separation would imply that transitions from one kind to the other do not occur. The preceding discussion is therefore only a rough one, applying to the case when such a separation is approximately possible.)

In this way we are led to infer that the negative-energy solutions of (48) refer to the motion of a new kind of particle having the mass of an electron and the opposite charge. Such particles have been observed experimentally and are called *positrons*. We cannot, however, simply assert that the negative-energy solutions represent positrons, as this would make the dynamical relations all wrong. For instance, it is certainly not true that a positron has a negative kinetic energy. We must therefore establish the theory of the positrons on a somewhat different footing. We assume that *nearly all the negative-energy states are occupied*, with one electron in each state in accordance with the exclusion principle of Pauli. An unoccupied negative-energy state will now appear as something with a positive energy, since to make it disappear, i.e. to fill it up, we should have to add to it an electron with negative energy. We assume that *these unoccupied negative-energy states are the positrons*.

These assumptions require there to be a distribution of electrons of infinite density everywhere in the world. A perfect vacuum is a region where all the states of positive energy are unoccupied and all those of negative energy are occupied. In a perfect vacuum Maxwell's equation

$$\operatorname{div} \mathcal{E} = 0$$

must, of course, be valid. This means that the infinite distribution of negative-energy electrons does not contribute to the electric field. Only departures from the distribution in a vacuum will contribute to the electric density  $\rho$  in Maxwell's equation

$$\operatorname{div} \mathcal{E} = 4\pi\rho.$$



Thus there will be a contribution  $-e$  for each occupied state of positive energy and a contribution  $+e$  for each unoccupied state of negative energy.

The exclusion principle will operate to prevent a positive-energy electron ordinarily from making transitions to states of negative energy. It will still be possible, however, for such an electron to drop into an unoccupied state of negative energy. In this case we should have an electron and positron disappearing simultaneously, their energy being emitted in the form of radiation. The converse process would consist in the creation of an electron and a positron from electromagnetic radiation.

The theory of the positron here given appears at first sight to treat the electrons and positrons on very different footings, but actually the fundamental ideas of the theory are symmetrical between the electrons and positrons. We should have an equivalent theory if we supposed the positrons to be the basic particles, described by wave equations of the form (9) with  $-e$  for  $e$ , and then supposed that nearly all the states of negative energy for the positron are filled up, a hole in the distribution of negative-energy positrons being then interpreted as an ordinary electron. The theory could be developed consistently with the hypothesis that all the laws of physics are symmetrical between positive and negative electric charge.



### XIII

#### FIELD THEORY

##### 76. Quantum Conditions for the Electromagnetic Field

THE methods of classical mechanics can be applied, not only to particles, but to the vibrations of a field such as the electromagnetic field. One can introduce dynamical variables to describe the field and can set up a Hamiltonian function which enables the equations of motion to be expressed in the Hamiltonian form. There exists a corresponding quantum mechanics of fields. It is of interest chiefly because of the mathematical beauty of its formal analogy with the classical theory when it is expressed in symbolic form. It has not so far led to any practical results which could not be obtained by more elementary methods.

We shall here deal with the quantum theory of the electromagnetic field. The foundations of this theory have already been given in Chapter XI, where we resolved the field into plane waves and treated the amplitudes and phases of these waves as dynamical variables. The present theory will go beyond that of Chapter XI in that the field quantities themselves will be used as dynamical variables, not merely the amplitudes and phases of their Fourier components, and the whole of the mutual interaction between electrified particles, including also the Coulomb interaction, will be shown to follow from the interaction between the particles and the field. The present theory will be relativistic throughout and we shall take the velocity of light  $c$  equal to unity.

We shall work for the present with the Heisenberg picture of § 32, in order that we may have our dynamical variables satisfying equations of motion analogous to those of classical mechanics. When we use the field quantities as dynamical variables, the first problem that presents itself is to obtain their quantum conditions, a problem which was first solved by Jordan and Pauli.<sup>†</sup> The general solution of this problem would require us to obtain the commutability relation connecting any two field quantities at any two points of space-time  $\mathbf{x}', t'$  and  $\mathbf{x}'', t''$ . For building up a dynamical theory, though, it is sufficient to obtain the commutability relations connecting all the field quantities at one instant of time, corresponding to the fact that in our particle dynamics we had to obtain only those quantum

<sup>†</sup> Jordan and Pauli, *Z. f. Physik*, **47** (1927), 151.



conditions connecting the dynamical variables at one time. The more general commutability relations would then be determined by these, together with the equations of motion. From general grounds we should expect two field quantities at two points in space-time, neither of which lies inside or on the light-cone from the other, to commute, since one can measure either of these quantities without disturbing the other, on account of the velocity of propagation of the disturbance being limited by the requirements of relativity to the velocity of light. When we work out the P.B.'s connecting the field quantities at various points of space at the same time, we shall find, in agreement with the above expectation, that two field quantities commute unless they are at two points infinitely close to one another.

We may assume that the commutability relations connecting the field quantities at one instant of time are independent of whether there are charged particles present or not, since in our theory of particle dynamics we had the same quantum conditions for a system whether that system interacts with another system or not. Thus we may work with the case of no charged particles present, when the whole of the electromagnetic field can be resolved into plane waves, and use the form for  $\mathcal{E}$  and  $\mathcal{H}$  given by (36) of Chapter XI, namely, with  $t = 0$ ,

$$\left. \begin{aligned} \mathcal{E} &= \sum \int \mathcal{E}_a \cos[\gamma_a - (\mathbf{k}_a, \mathbf{x})] d\mathbf{k}_a \\ \mathcal{H} &= \sum \int \mathcal{H}_a \cos[\gamma_a - (\mathbf{k}_a, \mathbf{x})] d\mathbf{k}_a. \end{aligned} \right\} (1)$$

It is convenient to pass from the continuous range of values of  $\mathbf{k}_a$  to a discrete set, corresponding to the discrete set of photon states that we had in § 64, and so to replace the integrals in (1) by sums. We then get, with  $s_a$  having the same meaning as in § 64,

$$\begin{aligned} \mathcal{E} &= \sum_a \mathcal{E}_a \cos[\gamma_a - (\mathbf{k}_a, \mathbf{x})] s_a^{-1} \\ \mathcal{H} &= \sum_a \mathcal{H}_a \cos[\gamma_a - (\mathbf{k}_a, \mathbf{x})] s_a^{-1}. \end{aligned}$$

The length of the vector  $\mathcal{E}_a$  or  $\mathcal{H}_a$  is given by (38) and (32) of Chapter XI to be

$$|\mathcal{E}_a| = |\mathcal{H}_a| = \pi^{-1} (h\nu_a n_a s_a)^{\frac{1}{2}}. \quad (2)$$

Thus if we let  $\alpha_a$  and  $\beta_a$  denote vectors of unit length in the directions of  $\mathcal{E}_a$  and  $\mathcal{H}_a$  respectively, we have

$$\left. \begin{aligned} \mathcal{E} &= \pi^{-1} \sum_a (h\nu_a)^{\frac{1}{2}} \alpha_a n_a^{\frac{1}{2}} \cos[\gamma_a - (\mathbf{k}_a, \mathbf{x})] s_a^{-\frac{1}{2}} \\ \mathcal{H} &= \pi^{-1} \sum_a (h\nu_a)^{\frac{1}{2}} \beta_a n_a^{\frac{1}{2}} \cos[\gamma_a - (\mathbf{k}_a, \mathbf{x})] s_a^{-\frac{1}{2}}. \end{aligned} \right\} (3)$$



These expressions for  $\mathcal{E}$  and  $\mathcal{H}$  hold in the classical theory. To pass over to the quantum theory, we must consider the  $n$ 's and  $\gamma$ 's in (3) to be non-commuting variables satisfying quantum conditions like (10) of Chapter XI with  $\gamma$  instead of  $w$ . The expressions on the right-hand side of (3) are then no longer real. To make them real, we have to replace  $2n_a^{\frac{1}{2}} \cos[\gamma_a - (\mathbf{k}_a, \mathbf{x})]$  by

$$n_a^{\frac{1}{2}} e^{i\gamma_a} e^{-i(\mathbf{k}_a, \mathbf{x})} + e^{-i\gamma_a} n_a^{\frac{1}{2}} e^{i(\mathbf{k}_a, \mathbf{x})}$$

or by

$$e^{i\gamma_a} n_a^{\frac{1}{2}} e^{-i(\mathbf{k}_a, \mathbf{x})} + n_a^{\frac{1}{2}} e^{-i\gamma_a} e^{i(\mathbf{k}_a, \mathbf{x})}.$$

We choose the first of these alternatives since the second, in a representation in which the  $n$ 's are diagonal, would make all matrix elements of  $\mathcal{E}$  and  $\mathcal{H}$ , lying in a row or a column for which an  $n'_a = 0$ , vanish and thus 0 would cease to be an effective eigenvalue of the  $n$ 's. The second alternative would not give a different physical theory, but would merely mean working with variables  $n$  which are greater by one than the numbers of photons in the various states. Equations (3) now become

$$\begin{aligned} \mathcal{E} &= (2\pi)^{-1} \sum_a (h\nu_a)^{\frac{1}{2}} \alpha_a \{ n_a^{\frac{1}{2}} e^{i\gamma_a} e^{-i(\mathbf{k}_a, \mathbf{x})} + e^{-i\gamma_a} n_a^{\frac{1}{2}} e^{i(\mathbf{k}_a, \mathbf{x})} \} s_a^{-\frac{1}{2}} \\ &= (2\pi)^{-1} \sum_a (h\nu_a)^{\frac{1}{2}} \alpha_a \{ \bar{\xi}_a e^{-i(\mathbf{k}_a, \mathbf{x})} + \xi_a e^{i(\mathbf{k}_a, \mathbf{x})} \} s_a^{-\frac{1}{2}}, \end{aligned} \quad (4)$$

with the introduction of variables  $\xi_a$  and  $\bar{\xi}_a$  like those of equations (13), Chapter XI, and similarly,

$$\mathcal{H} = (2\pi)^{-1} \sum_a (h\nu_a)^{\frac{1}{2}} \beta_a \{ \bar{\xi}_a e^{-i(\mathbf{k}_a, \mathbf{x})} + \xi_a e^{i(\mathbf{k}_a, \mathbf{x})} \} s_a^{-\frac{1}{2}}. \quad (5)$$

Let  $\mathcal{E}_l$  be the component of  $\mathcal{E}$  in a certain direction  $l$  and  $\mathcal{E}_m$  the component of  $\mathcal{E}$  in another direction  $m$ , which may, as a special case, be the same as the direction  $l$ . We shall now work out the P.B. connecting  $\mathcal{E}_l$  at the point  $x', y', z'$ , which we write as  $\mathcal{E}_l^{(')}$ , with  $\mathcal{E}_m$  at the point  $x'', y'', z''$ , which we write  $\mathcal{E}_m^{(')}$ . Since  $\xi$ 's and  $\bar{\xi}$ 's with different suffixes always commute with each other, we obtain from (4)

$$\begin{aligned} &[\mathcal{E}_l^{(')}, \mathcal{E}_m^{(')}] \\ &= (2\pi)^{-2} \sum_a h\nu_a \alpha_{la} \alpha_{ma} [\bar{\xi}_a e^{-i(\mathbf{k}_a, \mathbf{x}')} + \xi_a e^{i(\mathbf{k}_a, \mathbf{x}')} \bar{\xi}_a e^{-i(\mathbf{k}_a, \mathbf{x}'')} + \xi_a e^{i(\mathbf{k}_a, \mathbf{x}')} \xi_a e^{i(\mathbf{k}_a, \mathbf{x}'')}] s_a^{-1} \\ &= (2\pi)^{-2} \sum_a h\nu_a \alpha_{la} \alpha_{ma} (i\hbar)^{-1} \{ e^{i(\mathbf{k}_a, \mathbf{x}' - \mathbf{x}'')} - e^{-i(\mathbf{k}_a, \mathbf{x}' - \mathbf{x}'')} \} s_a^{-1}, \end{aligned}$$

from the quantum condition

$$\xi_a \bar{\xi}_a - \bar{\xi}_a \xi_a = 1, \quad (6)$$



which comes from (15) of Chapter XI. This reduces to

$$[\mathcal{E}_l^{(')}, \mathcal{E}_m^{(')}'] = \pi^{-1} \sum_a \nu_a \alpha_{la} \alpha_{ma} \sin(\mathbf{k}_a, \mathbf{x}' - \mathbf{x}'') s_a^{-1}. \quad (7)$$

When we do the summation over all values of  $a$  here, each value of  $a$  meaning a direction of motion and frequency of a Fourier component of the field, together with a direction of polarization, we shall evidently get the result zero, since each term in the sum will just cancel with the term corresponding to the opposite direction of motion, the same frequency and the same direction of the electric vector. Hence *the electric field quantities at different places and the same time all commute with each other*. Similarly, working from (5), it may be shown that *the magnetic field quantities at different places and the same time all commute with each other*.

It remains for us to determine the P.B. connecting the electric field at one place with the magnetic field at another. By similar work to that which led to (7) we obtain

$$[\mathcal{E}_l^{(')}, \mathcal{H}_m^{(')}'] = \pi^{-1} \sum_a \nu_a \alpha_{la} \beta_{ma} \sin(\mathbf{k}_a, \mathbf{x}' - \mathbf{x}'') s_a^{-1}. \quad (8)$$

Let us first do the summation here with respect to both states of polarization for a given direction of motion and frequency. The state of polarization is specified by the two mutually perpendicular vectors  $\alpha$  and  $\beta$ , giving the directions of the electric and magnetic fields, and to pass from one state of polarization to the other we merely have to replace  $\alpha$  by  $\beta$  and  $\beta$  by  $-\alpha$ . Thus to sum over both states of polarization in (8), we have to replace  $\alpha_{la} \beta_{ma}$  by

$$\alpha_{la} \beta_{ma} - \beta_{la} \alpha_{ma}. \quad (9)$$

If the two directions  $l$  and  $m$  are the same, expression (9) vanishes and thus the right-hand side of (8) vanishes. Hence *components of  $\mathcal{E}$  and  $\mathcal{H}$  in parallel directions always commute*.

We shall now take the case when the directions  $l$  and  $m$  are perpendicular and suppose them, for definiteness, to be the  $x$  and  $y$  directions. Expression (9) then becomes

$$\alpha_{xa} \beta_{ya} - \beta_{xa} \alpha_{ya},$$

which is just the  $z$ -component of the vector product  $\alpha_a \times \beta_a$ . Since this vector product is of unit length and is in the direction of the vector  $\mathbf{k}_a$ , its  $z$ -component is the cosine of the angle between the vector  $\mathbf{k}_a$  and the  $z$ -axis and is thus  $k_{za}/2\pi\nu_a$ . Hence equation (8) becomes

$$[\mathcal{E}_x^{(')}, \mathcal{H}_y^{(')}'] = \frac{1}{2} \pi^{-2} \sum_{\mathbf{k}_a} k_{za} \sin(\mathbf{k}_a, \mathbf{x}' - \mathbf{x}'') s_a^{-1}.$$



Passing from the sum back to an integral, we get

$$\begin{aligned} [\mathcal{E}_x^{(')}, \mathcal{H}_y^{(')}'] &= \frac{1}{2}\pi^{-2} \int k_z \sin(\mathbf{k}, \mathbf{x}' - \mathbf{x}'') d\mathbf{k} \\ &= -\frac{1}{2}\pi^{-2} \frac{\partial}{\partial z'} \int \cos(\mathbf{k}, \mathbf{x}' - \mathbf{x}'') d\mathbf{k} \\ &= -4\pi \frac{\partial}{\partial z'} \{\delta(x' - x'') \delta(y' - y'') \delta(z' - z'')\} \end{aligned}$$

with the help of formula (15) of Chapter IV. Thus we obtain finally

$$\left. \begin{aligned} [\mathcal{E}_x^{(')}, \mathcal{H}_y^{(')}'] &= -4\pi \delta(x' - x'') \delta(y' - y'') \delta'(z' - z'') \\ \text{and similarly,} \\ [\mathcal{E}_x^{(')}, \mathcal{H}_z^{(')}'] &= 4\pi \delta(x' - x'') \delta'(y' - y'') \delta(z' - z''), \end{aligned} \right\} (10)$$

with corresponding relations for  $\mathcal{E}_y$  and  $\mathcal{E}_z$ . This gives us all the quantum conditions for the field quantities at a definite time. Two of these field quantities always commute if they are at two points in space which are not infinitely close.

The total energy of the field in the absence of any charged particles is

$$H_F = \sum_a n_a h \nu_a = h \sum_a \nu_a \bar{\xi}_a \xi_a, \quad (11)$$

from (14) of Chapter XI. From (4)

$$\begin{aligned} \int \mathcal{E}^2 d\mathbf{x} &= (2\pi)^{-2} h \sum_{ab} \nu_a^{\frac{1}{2}} \nu_b^{\frac{1}{2}} (\alpha_a, \alpha_b) \int \{\bar{\xi}_a e^{-i(\mathbf{k}_a, \mathbf{x})} + \xi_a e^{i(\mathbf{k}_a, \mathbf{x})}\} \times \\ &\quad \times \{\bar{\xi}_b e^{-i(\mathbf{k}_b, \mathbf{x})} + \xi_b e^{i(\mathbf{k}_b, \mathbf{x})}\} d\mathbf{x} s_a^{-\frac{1}{2}} s_b^{-\frac{1}{2}} \\ &= 2\pi h \sum_{ab} \nu_a^{\frac{1}{2}} \nu_b^{\frac{1}{2}} (\alpha_a, \alpha_b) \{(\bar{\xi}_a \bar{\xi}_b + \xi_a \xi_b) \delta(\mathbf{k}_a + \mathbf{k}_b) + \\ &\quad + (\bar{\xi}_a \xi_b + \xi_a \bar{\xi}_b) \delta(\mathbf{k}_a - \mathbf{k}_b)\} s_a^{-\frac{1}{2}} s_b^{-\frac{1}{2}}, \quad (12) \end{aligned}$$

the  $\delta$  function of a vector having the same meaning as in equation (19) of Chapter IX. Similarly,

$$\begin{aligned} \int \mathcal{H}^2 d\mathbf{x} &= 2\pi h \sum_{ab} \nu_a^{\frac{1}{2}} \nu_b^{\frac{1}{2}} (\beta_a, \beta_b) \{(\bar{\xi}_a \bar{\xi}_b + \xi_a \xi_b) \delta(\mathbf{k}_a + \mathbf{k}_b) + \\ &\quad + (\bar{\xi}_a \xi_b + \xi_a \bar{\xi}_b) \delta(\mathbf{k}_a - \mathbf{k}_b)\} s_a^{-\frac{1}{2}} s_b^{-\frac{1}{2}}. \end{aligned}$$

Hence

$$\begin{aligned} \int (\mathcal{E}^2 + \mathcal{H}^2) d\mathbf{x} &= 2\pi h \sum_{ab} \nu_a^{\frac{1}{2}} \nu_b^{\frac{1}{2}} \{(\alpha_a, \alpha_b) + (\beta_a, \beta_b)\} \times \\ &\quad \times (\bar{\xi}_a \xi_b + \xi_a \bar{\xi}_b) \delta(\mathbf{k}_a - \mathbf{k}_b) s_a^{-\frac{1}{2}} s_b^{-\frac{1}{2}}, \end{aligned}$$

the remaining terms cancelling since  $(\alpha_a, \alpha_b) = -(\beta_a, \beta_b)$  when the vectors  $\mathbf{k}_a$  and  $\mathbf{k}_b$  are in opposite directions. This result reduces to

$$\int (\mathcal{E}^2 + \mathcal{H}^2) d\mathbf{x} = 4\pi h \sum_a \nu_a (\bar{\xi}_a \xi_a + \xi_a \bar{\xi}_a),$$



since, as is easily verified,  $\delta(\mathbf{k}_a - \mathbf{k}_b) s_a^{-\frac{1}{2}} s_b^{-\frac{1}{2}}$  occurring in a sum over  $a$  or  $b$  is equivalent to  $\delta_{\mathbf{k}_a \mathbf{k}_b}$ . With the help of (6) we now find

$$\int (\mathcal{E}^2 + \mathcal{H}^2) d\mathbf{x} = 8\pi\hbar \sum_a \nu_a (\bar{\xi}_a \xi_a + \tfrac{1}{2}),$$

and hence the energy (11) becomes

$$H_F = 1/8\pi \cdot \int (\mathcal{E}^2 + \mathcal{H}^2) d\mathbf{x} - \tfrac{1}{2} \sum_a \hbar \nu_a. \quad (13)$$

Thus the classical expression for the energy of the electromagnetic field, given by (37) of Chapter XI, holds in the quantum theory only when another term is added to it, consisting of a contribution  $-\frac{1}{2}\hbar\nu$  for each degree of freedom of the field. This extra term is infinite, but it is a constant, independent of all the dynamical variables, and may therefore be neglected when (13) is used as a Hamiltonian.

The equations of motion of the field may be deduced directly from the quantum conditions for the field quantities and the form (13) for the Hamiltonian, without any resolution of the field into Fourier components. Thus, for example, we get as the equation of motion for  $\mathcal{E}_x^{(')}$

$$\dot{\mathcal{E}}_x^{(')} = [\mathcal{E}_x^{(')}, H_F] = 1/8\pi \cdot \int [\mathcal{E}_x^{(')}, \mathcal{H}_y^{(n)2} + \mathcal{H}_z^{(n)2}] d\mathbf{x}'' \quad (14)$$

Now from (10)

$$\begin{aligned} [\mathcal{E}_x^{(')}, \mathcal{H}_y^{(n)2}] &= [\mathcal{E}_x^{(')}, \mathcal{H}_y^{(n)}] \mathcal{H}_y^{(n)} + \mathcal{H}_y^{(n)} [\mathcal{E}_x^{(')}, \mathcal{H}_y^{(n)}] \\ &= -8\pi \mathcal{H}_y^{(n)} \delta(x' - x'') \delta(y' - y'') \delta'(z' - z''), \end{aligned}$$

and hence

$$\begin{aligned} \int [\mathcal{E}_x^{(')}, \mathcal{H}_y^{(n)2}] d\mathbf{x}'' &= -8\pi \int \int \int \mathcal{H}_y^{(n)} \delta(x' - x'') \delta(y' - y'') \delta'(z' - z'') dx'' dy'' dz'' \\ &= -8\pi \frac{\partial}{\partial z'} \mathcal{H}_y^{(n)}, \end{aligned}$$

from (4) and (6) of Chapter IV. Similarly,

$$\int [\mathcal{E}_x^{(')}, \mathcal{H}_z^{(n)2}] d\mathbf{x}'' = 8\pi \frac{\partial}{\partial y'} \mathcal{H}_z^{(n)}.$$

Thus (14) reduces to

$$\dot{\mathcal{E}}_x^{(')} = -\frac{\partial}{\partial z'} \mathcal{H}_y^{(n)} + \frac{\partial}{\partial y'} \mathcal{H}_z^{(n)}, \quad (15)$$

which is one of Maxwell's equations.

## 77. Quantum Conditions for the Electromagnetic Potentials

The foregoing theory of the quantum conditions for the electromagnetic field quantities  $\mathcal{E}$  and  $\mathcal{H}$  in a vacuum must now be extended to include the quantum conditions for the potentials  $A_0$  and  $\mathbf{A}$ . It



might be thought that such an extension is not necessary, since only the field quantities  $\mathcal{E}$  and  $\mathcal{H}$  are physically significant. The potentials appear, however, in the equations describing the interaction of a charged particle with the field, so that when we come to take the presence of charged particles into account, as we shall do later, we shall need to know all about the potentials.

The problem of including the potentials in our theory is not a straightforward one, owing to the potentials not being uniquely determined in terms of the field quantities  $\mathcal{E}$  and  $\mathcal{H}$ . The arbitrariness in the potentials can be reduced by imposing on them the condition that their four-dimensional divergence shall vanish, i.e.

$$\partial A_0 / \partial t + \text{div } \mathbf{A} = 0, \quad (16)$$

but even then they are not completely determined. In Chapter XI we made them definite by taking  $A_0 = 0$ , but such an assumption is not relativistic and so cannot be made here. For the present we shall ignore equation (16) and the arbitrariness in the potentials and shall return to these questions in the next section.

We express the potentials in terms of their Fourier components, like we did the field quantities  $\mathcal{E}$  and  $\mathcal{H}$  in (1), thus

$$A_\mu = \int A_{\mu\mathbf{k}} \cos[\gamma_{\mathbf{k}} + 2\pi\nu_{\mathbf{k}}t - (\mathbf{k}, \mathbf{x})] d\mathbf{k}, \quad (17)$$

where the suffix  $\mu$  takes on any of the values 0,  $x$ ,  $y$ ,  $z$ . It is necessary to do this Fourier resolution for all time, and not merely for the time  $t = 0$ , as we did in (1), owing to the fact that the potentials are not determined throughout all time if they are given at one time, as is the case with  $\mathcal{E}$  and  $\mathcal{H}$ . The amplitudes  $A_{\mu\mathbf{k}}$  in (17) are specified by the suffix  $\mathbf{k}$ , and not by a suffix  $a$  as in (1), since there is only one of them corresponding to any value of  $\mathbf{k}$ , and not two, referring to two different states of polarization, as in (1). We again pass from integrals to sums and get, corresponding to the forms (4) and (5) for  $\mathcal{E}$  and  $\mathcal{H}$ ,

$$\begin{aligned} A_\mu &= \sum_{\mathbf{k}} A_{\mu\mathbf{k}} \cos[\gamma_{\mathbf{k}} + 2\pi\nu_{\mathbf{k}}t - (\mathbf{k}, \mathbf{x})] s_{\mathbf{k}}^{-1} \\ &= \sum_{\mathbf{k}} \{ \bar{\eta}_{\mu\mathbf{k}} e^{i[2\pi\nu_{\mathbf{k}}t - (\mathbf{k}, \mathbf{x})]} + \eta_{\mu\mathbf{k}} e^{-i[2\pi\nu_{\mathbf{k}}t - (\mathbf{k}, \mathbf{x})]} \} s_{\mathbf{k}}^{-\frac{1}{2}}, \end{aligned} \quad (18)$$

with the variables  $\eta$  playing roughly the part of the  $\xi$ 's in (4) and (5). We may write this result as

$$A_\mu = \sum_{\mathbf{k}} \{ \bar{\zeta}_{\mu\mathbf{k}} e^{-i(\mathbf{k}, \mathbf{x})} + \zeta_{\mu\mathbf{k}} e^{i(\mathbf{k}, \mathbf{x})} \} s_{\mathbf{k}}^{-\frac{1}{2}}, \quad (19)$$

where

$$\zeta_{\mu\mathbf{k}} = \eta_{\mu\mathbf{k}} e^{-2\pi i \nu_{\mathbf{k}} t}, \quad \bar{\zeta}_{\mu\mathbf{k}} = \bar{\eta}_{\mu\mathbf{k}} e^{2\pi i \nu_{\mathbf{k}} t}. \quad (20)$$



The  $\zeta$ 's and  $\bar{\zeta}$ 's are dynamical variables not involving the time explicitly, since the equations connecting them with the dynamical variables  $A_\mu$  do not involve the time explicitly. They thus satisfy equations of motion of the form (10) of Chapter VI, rather than (13) of Chapter VI. These equations of motion, when there are no charged particles present, must be such as to make the  $\zeta$ 's and  $\bar{\zeta}$ 's vary with time according to equations (20) with the  $\eta$ 's and  $\bar{\eta}$ 's constants.

We must now obtain the quantum conditions for the  $\zeta$ 's and  $\bar{\zeta}$ 's at a particular time. Let us study first the  $\zeta$ 's and  $\bar{\zeta}$ 's describing one particular Fourier component of the field, consisting of waves moving in the direction of the  $x$ -axis with a definite frequency  $\nu$ , so that we have  $k_y = k_z = 0$ ,  $k_x = 2\pi\nu$ . We shall then have  $\zeta_y$  and  $\zeta_z$  determined by the field quantities  $\mathcal{E}$  and  $\mathcal{H}$ . According to the equations

$$\mathcal{E}_y = -\frac{\partial A_y}{\partial t} - \frac{\partial A_0}{\partial y}, \quad \mathcal{E}_z = -\frac{\partial A_z}{\partial t} - \frac{\partial A_0}{\partial z},$$

$\zeta_y$  and  $\zeta_z$  must each be  $(2\pi i\nu)^{-1}$  times that Fourier coefficient  $(2\pi)^{-1}(\hbar\nu)^{\frac{1}{2}}\xi_a$  in (4) belonging to  $k_y = k_z = 0$ ,  $k_x = 2\pi\nu$ , and to the  $y$  and  $z$  directions for the electric vector respectively. Thus from the quantum condition (6)

$$\left. \begin{aligned} \zeta_y \bar{\zeta}_y - \bar{\zeta}_y \zeta_y &= \hbar/16\pi^4\nu \\ \zeta_z \bar{\zeta}_z - \bar{\zeta}_z \zeta_z &= \hbar/16\pi^4\nu. \end{aligned} \right\} (21)$$

Further,  $\zeta_y$  and  $\bar{\zeta}_y$  commute with  $\zeta_z$  and  $\bar{\zeta}_z$ , since they belong to different degrees of freedom.

The  $\zeta_x$  and  $\zeta_0$  variables are not determined by  $\mathcal{E}$  and  $\mathcal{H}$ , so that we cannot deduce their quantum conditions in the same way. We have to make some new assumptions for them. To make equations (21) into a complete relativistic set, we assume

$$\left. \begin{aligned} \zeta_x \bar{\zeta}_x - \bar{\zeta}_x \zeta_x &= \hbar/16\pi^4\nu \\ \zeta_0 \bar{\zeta}_0 - \bar{\zeta}_0 \zeta_0 &= -\hbar/16\pi^4\nu. \end{aligned} \right\} (22)$$

The minus sign on the right-hand side of the second of these equations is required by relativity. Finally we assume generally

$$[\zeta_\mu, \zeta_\nu] = 0 \quad [\bar{\zeta}_\mu, \bar{\zeta}_\nu] = 0 \quad [\bar{\zeta}_\mu, \zeta_\nu] = 0 \quad \text{for } \mu \neq \nu. \quad (23)$$

This gives us the complete set of quantum conditions for the particular Fourier component we are considering. Those for the other Fourier components will be of similar form. Variables belonging to different Fourier components must, of course, commute.

We can now work out the quantum conditions connecting the



potentials at different points in space-time. Denoting by  $A_{\mu}^{(\prime)}(t')$  the value of  $A_{\mu}$  at the point  $\mathbf{x}', t'$ , we have in the first place, from (23),

$$[A_{\mu}^{(\prime)}(t'), A_{\nu}^{(\prime)}(t'')] = 0, \quad \mu \neq \nu. \quad (24)$$

Further, by the same kind of work as led to (7) or (8), we obtain

$$\begin{aligned} [A_{\mu}^{(\prime)}(t'), A_{\mu}^{(\prime)}(t'')] &= \pm \sum_{\mathbf{k}} 1/4\pi^3 \nu_{\mathbf{k}} \cdot \sin[(\mathbf{k}, \mathbf{x}' - \mathbf{x}'') - 2\pi\nu_{\mathbf{k}}(t' - t'')] s_{\mathbf{k}}^{-1} \\ &= \pm \int 1/4\pi^3 \nu_{\mathbf{k}} \cdot \sin[(\mathbf{k}, \mathbf{x}' - \mathbf{x}'') - 2\pi\nu_{\mathbf{k}}(t' - t'')] d\mathbf{k}, \end{aligned} \quad (25)$$

the upper sign being taken for  $\mu = x, y$ , or  $z$  and the lower sign for  $\mu = 0$ . The evaluation of the integral here leads to the result

$$\begin{aligned} [A_{\mu}^{(\prime)}(t'), A_{\mu}^{(\prime)}(t'')] &= \pm 4\pi \int \int \nu \sin[2\pi\nu(|\mathbf{x}' - \mathbf{x}''| \cos \theta - t' + t'')] d\nu d \cos \theta \\ &= \pm 2/|\mathbf{x}' - \mathbf{x}''| \cdot \int_0^{\infty} \{\cos[2\pi\nu(|\mathbf{x}' - \mathbf{x}''| + t' - t'')] - \\ &\quad - \cos[2\pi\nu(|\mathbf{x}' - \mathbf{x}''| - t' + t'')]\} d\nu \\ &= \pm 1/|\mathbf{x}' - \mathbf{x}''| \cdot \{\delta(|\mathbf{x}' - \mathbf{x}''| + t' - t'') - \delta(|\mathbf{x}' - \mathbf{x}''| - t' + t'')\}, \end{aligned} \quad (26)$$

from (15) of Chapter IV. The expression (26) vanishes when  $t' = t''$ , so that the potentials at a given time instant all commute with each other, that is

$$[A_{\mu}^{(\prime)}, A_{\mu}^{(\prime)}] = 0. \quad (27)$$

For  $t' < t''$ , the second term in the  $\{ \}$  brackets in (26) vanishes, so that we can change its sign. We then get, from (12) of Chapter IV,

$$[A_{\mu}^{(\prime)}(t'), A_{\mu}^{(\prime)}(t'')] = \pm 2\delta\{(\mathbf{x}' - \mathbf{x}'')^2 - (t' - t'')^2\}. \quad (28)$$

Similarly, for  $t' > t''$ ,

$$[A_{\mu}^{(\prime)}(t'), A_{\mu}^{(\prime)}(t'')] = \mp 2\delta\{(\mathbf{x}' - \mathbf{x}'')^2 - (t' - t'')^2\}. \quad (29)$$

This gives us all the quantum conditions for the potentials.

The useful quantum conditions, in connexion with the dynamical equations, are those connecting two  $A$ 's at different places at the same time, namely (27), those connecting an  $A$  and  $\partial A/\partial t$  at different places at the same time, and those connecting two  $\partial A/\partial t$ 's at different places at the same time, all these being required since the  $A$ 's and the  $\partial A/\partial t$ 's at a given time are independent. Those involving  $\partial A/\partial t$ 's may be obtained from (28) and (29), but can be obtained more easily from (25), since the  $\delta$  function in (28) and (29) is an awkward one to



differentiate. Differentiating (25) with respect to  $t'$  and putting  $t' = t'' = t$ , we obtain

$$\begin{aligned} [\partial A_{\mu}^{(\prime)}/\partial t, A_{\mu}^{(\prime\prime)}] &= \mp 1/2\pi^2 \cdot \int \cos(\mathbf{k}, \mathbf{x}' - \mathbf{x}'') d\mathbf{k} \\ &= \mp 4\pi\delta(\mathbf{x}' - \mathbf{x}''). \end{aligned} \quad (30)$$

Again, differentiating (25) with respect to  $t'$  and  $t''$  and putting  $t' = t'' = t$ , we obtain

$$[\partial A_{\mu}^{(\prime)}/\partial t, \partial A_{\mu}^{(\prime\prime)}/\partial t] = \mp 1/\pi \cdot \int v_{\mathbf{k}} \sin(\mathbf{k}, \mathbf{x}' - \mathbf{x}'') d\mathbf{k} = 0. \quad (31)$$

Let us now suppose the Hamiltonian (11) to be expressed in terms of the  $\zeta$  variables. The contribution to it of the particular Fourier component which we considered above, consisting of waves moving in the direction of the  $x$ -axis, will be

$$16\pi^4 v^2 (\bar{\zeta}_y \zeta_y + \bar{\zeta}_z \zeta_z), \quad (32)$$

as may be seen by referring to (11) and to the connexion between the  $\zeta$ 's and the  $\xi$ 's which we had in deriving (21). This contribution to the Hamiltonian will make  $\zeta_y$  and  $\zeta_z$  vary with time in the desired way, namely according to (20), but it will make  $\zeta_x$  and  $\zeta_0$  constants of the motion, since they commute with (32). It therefore becomes necessary to modify the Hamiltonian and to replace the contribution (32) by

$$16\pi^4 v^2 (\bar{\zeta}_x \zeta_x + \bar{\zeta}_y \zeta_y + \bar{\zeta}_z \zeta_z - \zeta_0 \bar{\zeta}_0), \quad (33)$$

in order that all four  $\zeta$ 's may vary with time according to (20). It is better to put the  $\zeta_0$  to the left of the  $\bar{\zeta}_0$ , as will be seen in the next section, equation (44). With (33) in the Hamiltonian,  $\zeta_x$ ,  $\zeta_y$  and  $\zeta_z$  may be pictured as describing three harmonic oscillators of the ordinary kind, and  $\zeta_0$  a fourth harmonic oscillator of negative mass. The total Hamiltonian is now

$$H_F = 16\pi^4 \sum_{\mathbf{k}} v_{\mathbf{k}}^2 (\bar{\zeta}_{x\mathbf{k}} \zeta_{x\mathbf{k}} + \bar{\zeta}_{y\mathbf{k}} \zeta_{y\mathbf{k}} + \bar{\zeta}_{z\mathbf{k}} \zeta_{z\mathbf{k}} - \zeta_{0\mathbf{k}} \bar{\zeta}_{0\mathbf{k}}). \quad (34)$$

The physical effect of the extra terms that have been introduced here will be discussed in the next section.

It may be noted that equations (20), giving the integrals of the equations of motion for the  $\zeta$ 's and  $\bar{\zeta}$ 's, must be equivalent to

$$\zeta_{\mu\mathbf{k}} = e^{iH_F t/\hbar} \eta_{\mu\mathbf{k}} e^{-iH_F t/\hbar}, \quad (35)$$

from (18) of Chapter VI.

## 78. The Supplementary Conditions

We must now consider what we are to do with the classical equation (16). We cannot take it over directly into the quantum theory with-



out getting inconsistencies. For example, the P.B. of the left-hand side of (16) with  $A_0^{(v)}$  does not vanish, according to the quantum conditions (24) and (30), and so this left-hand side itself cannot vanish. The way out of the difficulty was shown by Fermi.<sup>†</sup> It consists in adopting a less stringent equation than (16), namely the equation

$$\{\partial A_0/\partial t + \text{div } \mathbf{A}\}\psi = 0, \quad (36)$$

and assuming it to hold for any  $\psi$  representing a state that can actually occur in nature. The operator in (36) involves  $\mathbf{x}$  and  $t$  as parameters, so there is one equation (36) for each set of values for  $\mathbf{x}$  and  $t$ , and these equations must all hold for any  $\psi$  representing a state that can actually occur. [The  $\psi$  in (36) does not depend on  $t$ , since we are using the Heisenberg picture, in which each state is represented by a fixed  $\psi$ .]

We shall call a condition, such as (36), which a  $\psi$  has to satisfy to represent an actual state, a *supplementary condition*. The existence of supplementary conditions in our theory does not mean any departure from or modification in the general principles of quantum mechanics. The principle of superposition of states and the whole of the general theory of states and observables, as given in Chapter II, apply also when there are supplementary conditions, provided we impose a further requirement on a linear operator in order that it may represent an observable, namely the requirement that, when it operates on any  $\psi$  satisfying the supplementary conditions, it changes this  $\psi$  into another  $\psi$  satisfying the supplementary conditions. We have already had an example of supplementary conditions in the theory of systems containing several similar particles. The condition that only symmetrical wave functions, or only antisymmetrical wave functions, represent states that can actually occur in nature, is precisely of the same type as condition (36) and is what we are now calling a supplementary condition. In this theory the further requirement on linear operators in order that they shall represent observables is that they shall be symmetrical between the similar particles.

When we introduce supplementary conditions into our theory we must verify that they are not too restrictive to allow any  $\psi$  at all to satisfy them. If we have more than one supplementary condition, we can deduce further supplementary conditions from them by taking P.B.'s of the operators in them; thus if we have

$$U\psi = 0, \quad V\psi = 0, \quad (37)$$

<sup>†</sup> Fermi, *Reviews of Modern Physics*, 4 (1932), 125.



we can deduce

$$[U, V]\psi = 0, \quad [U, [U, V]]\psi = 0, \quad (38)$$

and so on. To verify that our supplementary conditions are not too restrictive, we have to look into all the further supplementary conditions obtainable by this procedure to see that they can be satisfied, which we can usually do by showing that after a certain point the further supplementary conditions are all either identically satisfied or repetitions of the previous ones.

Since the left-hand side of (36) must vanish for all values of  $\mathbf{x}$ ,  $t$ , we can resolve it into its Fourier components and each component must vanish separately. From (18) and (20) this gives us equations of the form

$$\left. \begin{aligned} (\zeta_0 - \zeta_x)\psi &= 0 \\ (\bar{\zeta}_0 - \bar{\zeta}_x)\psi &= 0, \end{aligned} \right\} \quad (39)$$

where we have taken as a typical case the Fourier component we had in the preceding section referring to waves moving in the direction of the  $x$ -axis. Equations (39) are equivalent to (36) and are more convenient to work with for many purposes. The two operators in the two equations (39) we have written down commute with each other on account of (22) and (23). It follows that all the operators in all the various equations (39) referring to the various Fourier components will commute with each other, since variables belonging to different Fourier components commute. Hence our supplementary conditions are not too restrictive, all the further conditions obtainable from them in the way (38) was obtained from (37) being identically satisfied. The effect of the supplementary conditions (39) is to stop the  $\zeta_0$  and  $\zeta_x$  variables from contributing to the number of degrees of freedom, so that we are left with only two degrees of freedom for each frequency and direction of motion.

Since equation (16) is not valid and has to be replaced by a supplementary condition, any consequences of (16) in the ordinary Maxwell theory will not be valid in the quantum theory and will have to be replaced by supplementary conditions. The equations

$$\operatorname{div} \mathcal{H} = 0, \quad \partial \mathcal{H} / \partial t = -\operatorname{curl} \mathcal{E} \quad (40)$$

follow simply from the equations defining  $\mathcal{E}$  and  $\mathcal{H}$ ,

$$\mathcal{E} = -\partial \mathbf{A} / \partial t - \operatorname{grad} A_0, \quad \mathcal{H} = \operatorname{curl} \mathbf{A}, \quad (41)$$

and are therefore valid also in the quantum theory. The other Maxwell equations for empty space, however, namely

$$\operatorname{div} \mathcal{E} = 0, \quad \partial \mathcal{E} / \partial t = \operatorname{curl} \mathcal{H}, \quad (42)$$



can be derived only with the help of (16), and are thus not valid in the quantum theory. They must be replaced by

$$\{\text{div } \mathcal{E}\}\psi = 0, \quad \{\partial \mathcal{E}/\partial t - \text{curl } \mathcal{H}\}\psi = 0, \quad (43)$$

holding for any  $\psi$  representing a state that can actually occur. The failure of the second of equations (42) is connected with the change which we must make in the Hamiltonian from what it was when we derived (15).

The extra terms which we introduced in passing from (32) to (33) are

$$16\pi^4\nu^2(\bar{\zeta}_x\zeta_x - \zeta_0\bar{\zeta}_0) = 8\pi^4\nu^2\{(\bar{\zeta}_x + \bar{\zeta}_0)(\zeta_x - \zeta_0) + (\zeta_x + \zeta_0)(\bar{\zeta}_x - \bar{\zeta}_0)\}, \quad (44)$$

from (22) and (23). Thus these extra terms vanish when multiplied into any  $\psi$  satisfying (39). Hence the total change in the Hamiltonian, i.e. the difference between (34) and (11), will vanish when multiplied into any  $\psi$  representing a state that can actually occur and will therefore be physically unobservable. The new Hamiltonian may be put in the form

$$H_F = 1/8\pi \cdot \int \sum_{\mu} \pm \{(\text{grad } A_{\mu}, \text{grad } A_{\mu}) + (\partial A_{\mu}/\partial t)^2\} \mathbf{dx} - \frac{1}{2} \sum_a h\nu_a, \quad (45)$$

where the  $+$  sign is to be taken for  $\mu = x, y, z$  and the  $-$  sign for  $\mu = 0$ , by the same kind of analysis as that by which the old one was put in the form (13). This new Hamiltonian will lead, of course, to the correct equations of motion for the  $A_{\mu}$  and  $\partial A_{\mu}/\partial t$ , namely

$$[A_{\mu}, H_F] = \partial A_{\mu}/\partial t \quad (46)$$

$$[\partial A_{\mu}/\partial t, H_F] = \nabla^2 A_{\mu}, \quad (47)$$

as may easily be verified from the quantum conditions (24), (27), (30), and (31).

It should be noticed that the field quantities  $\mathcal{E}$  and  $\mathcal{H}$  commute with the operators in the supplementary conditions. This follows from the fact that, if we take a Fourier component referring to waves moving in the direction of the  $x$ -axis, from (41) its contributions to  $\mathcal{E}_y$ ,  $\mathcal{E}_z$ ,  $\mathcal{H}_y$  and  $\mathcal{H}_z$  will depend only on  $\zeta_y$ ,  $\zeta_z$ ,  $\bar{\zeta}_y$  and  $\bar{\zeta}_z$ , its contribution to  $\mathcal{H}_x$  will vanish, and its contribution to  $\mathcal{E}_x$  will depend only on  $(\zeta_x - \zeta_0)$  and  $(\bar{\zeta}_x - \bar{\zeta}_0)$ . Each of these  $\zeta$ 's or  $\bar{\zeta}$ 's or combination of  $\zeta$ 's or  $\bar{\zeta}$ 's commutes with the operators in (39). From this commuting of  $\mathcal{E}$  and  $\mathcal{H}$  with the operators in the supplementary conditions we can infer that, when  $\mathcal{E}$  or  $\mathcal{H}$  is multiplied into a  $\psi$  satisfying the supplementary conditions, it will give another  $\psi$  satisfying the supplementary conditions, and hence it fulfils the new requirement for being



an observable. Further, the field energy  $H_F$  is composed of terms which are functions of  $\mathcal{E}$  and  $\mathcal{H}$  and terms which vanish when multiplied into a  $\psi$  satisfying the supplementary conditions. Hence  $H_F$  multiplied into a  $\psi$  satisfying the supplementary conditions gives another  $\psi$  satisfying the supplementary conditions, and thus  $H_F$  fulfils the new requirement for being an observable.

## 79. Interaction of Field and Particles

We shall now consider how the presence of charged particles in the field is to be taken into account—a problem that was first solved by Heisenberg and Pauli.<sup>†</sup> We can attack this problem by passing from the Heisenberg picture, which we have used exclusively in the three preceding sections, to the Schrödinger picture and setting up the Schrödinger wave equation with the Hamiltonian

$$H = H_F + \sum_r H_r, \quad (48)$$

where  $H_F$  is the energy of the field alone, given by (34) or (45), and  $H_r$  is the energy of the  $r$ th particle in interaction with the field. If we assume the particles are described by wave equations of the form of the relativistic wave equation for the electron, equation (9) of Chapter XII, we should have for  $H_r$ ,

$$H_r = e_r A_{0r} - (\alpha_r \mathbf{p}_r - e_r \mathbf{A}_r) \cdot \boldsymbol{\alpha}_{mr} m_r, \quad (49)$$

where  $e_r$  and  $m_r$  are the charge and mass of the  $r$ th particle,  $\mathbf{p}_r$  and the  $\alpha_r$ 's are dynamical variables describing this particle, and  $A_{0r}, \mathbf{A}_r$  are the potentials at the point where this particle is situated. These potentials are of the form (19), where the  $\zeta$ 's and  $\bar{\zeta}$ 's are now (like all dynamical variables in the Schrödinger picture) fixed operators, but satisfy the same quantum conditions as before.

We have as wave equation

$$i\hbar \frac{d\psi}{dt} = \{H_F + \sum_r H_r\} \psi. \quad (50)$$

This wave equation is not at all relativistic in its form, since it involves only one time variable  $t$ , but many sets of space variables  $x, y, z$ , one set for each particle. In order to get it into a relativistic form, it is necessary to introduce several time variables  $t_1, t_2, \dots, t_r, \dots$ , one for each particle. This can best be effected with the help of a

<sup>†</sup> Heisenberg and Pauli, *Z. f. Physik*, **56** (1929), 1; **59** (1929) 168.



certain contact transformation, by a method due to Rosenfeld.† We make the contact transformation of dynamical variables

$$\beta^* = e^{iH_F t/\hbar} \beta e^{-iH_F t/\hbar} \quad (51)$$

and put

$$\psi^* = e^{iH_F t/\hbar} \psi.$$

We then get, as the wave equation for  $\psi^*$ ,

$$\begin{aligned} i\hbar \frac{d\psi^*}{dt} &= e^{iH_F t/\hbar} \left\{ i\hbar \frac{d\psi}{dt} - H_F \psi \right\} \\ &= e^{iH_F t/\hbar} \sum_r H_r \psi \\ &= \sum_r H_r^* \psi^*. \end{aligned} \quad (52)$$

[This work, it may be noted, is essentially equivalent to that leading to equation (14) of Chapter VIII.] The wave equation (52) for  $\psi^*$  differs from the previous one (50) for  $\psi$  through the disappearance of  $H_F$  and the replacement of the  $H_r$ 's by  $H_r^*$ 's.

Let us now examine  $H_r^*$ . From (49) and (51) we obtain

$$\begin{aligned} H_r^* &= e^{iH_F t/\hbar} \{ e_r A_{0r} - (\alpha_r, \mathbf{p}_r - e_r \mathbf{A}_r) - \alpha_{mr} m_r \} e^{-iH_F t/\hbar} \\ &= e_r A_{0r}^* - (\alpha_r, \mathbf{p}_r - e_r \mathbf{A}_r^*) - \alpha_{mr} m_r, \end{aligned} \quad (53)$$

since  $H_F$  commutes with the  $p$ 's and with the  $\alpha$ 's. Further, from (19), with  $\mathbf{x}_r$  denoting the position of the  $r$ th particle,

$$A_{\mu r}^* = \sum_{\mathbf{k}} \{ \bar{\zeta}_{\mu \mathbf{k}}^* e^{-i(\mathbf{k}, \mathbf{x}_r)} + \zeta_{\mu \mathbf{k}}^* e^{i(\mathbf{k}, \mathbf{x}_r)} \} s_{\mathbf{k}}^{-\frac{1}{2}}$$

where

$$\zeta_{\mu \mathbf{k}}^* = e^{iH_F t/\hbar} \zeta_{\mu \mathbf{k}} e^{-iH_F t/\hbar}.$$

Now  $\zeta_{\mu \mathbf{k}}$  is a constant operator and is thus like the  $\eta_{\mu \mathbf{k}}$  of the preceding section, so that, from (35),  $\zeta_{\mu \mathbf{k}}^*$  must vary with  $t$  in the same way as the  $\zeta_{\mu \mathbf{k}}$  of the preceding section did, i.e. according to the law (20). Thus

$$\zeta_{\mu \mathbf{k}}^* = \zeta_{\mu \mathbf{k}} e^{-2\pi i \nu_{\mathbf{k}} t}, \quad (54)$$

so that

$$A_{\mu r}^* = \sum_{\mathbf{k}} \{ \bar{\zeta}_{\mu \mathbf{k}} e^{i[2\pi \nu_{\mathbf{k}} t - (\mathbf{k}, \mathbf{x}_r)]} + \zeta_{\mu \mathbf{k}} e^{-i[2\pi \nu_{\mathbf{k}} t - (\mathbf{k}, \mathbf{x}_r)]} \} s_{\mathbf{k}}^{-\frac{1}{2}}. \quad (55)$$

Equations (53) and (55) give us  $H_r^*$ . They show that  $H_r^*$  is of relativistic form in the space-time variables  $\mathbf{x}_r, t$ . We shall later require to use formula (55) applied to a general point  $\mathbf{x}$ , not necessarily one where a particle is situated, when it will read

$$A_{\mu}^* = \sum_{\mathbf{k}} \{ \bar{\zeta}_{\mu \mathbf{k}} e^{i[2\pi \nu_{\mathbf{k}} t - (\mathbf{k}, \mathbf{x})]} + \zeta_{\mu \mathbf{k}} e^{-i[2\pi \nu_{\mathbf{k}} t - (\mathbf{k}, \mathbf{x})]} \} s_{\mathbf{k}}^{-\frac{1}{2}}. \quad (56)$$

These  $A_{\mu}^*$ 's for various values of  $\mathbf{x}$  and  $t$  are similar to, and satisfy the same commutability relations as, the  $A_{\mu}$ 's in the Heisenberg picture with no charged particles present.

† Rosenfeld, *Z. f. Physik*, **76** (1932) 729.



We now introduce a wave function  $\Psi$  which involves, not just a single time variable  $t$ , like  $\psi^*$  does, but a whole set of them  $t_r$ , one for each particle, and suppose it satisfies the following set of wave equations

$$i\hbar \frac{\partial \Psi}{\partial t_r} = H_r^*(t_r) \Psi, \quad (57)$$

where  $H_r^*(t_r)$  is the  $H_r^*$  given by (53) and (55) with  $t_r$  substituted for  $t$ . There is one of these wave equations for each particle. These wave equations are obviously of relativistic form and we assume them as the fundamental equations describing the relativistic interaction of several charged particles with the field.

In order to justify the replacement of the wave equation (52) by the wave equations (57), we ought to verify, firstly, that when we put all the time variables  $t_r$  in  $\Psi$  equal to  $t$ , we get a  $\psi^*$  which satisfies (52), and secondly, that every  $\psi^*$  satisfying (52) can be generalized to a  $\Psi$  satisfying (57). The first of these conditions follows at once from the fact that the operator  $d/dt$  applied after we have put all the  $t_r$  equal to  $t$  is equivalent to  $\sum_r \partial/\partial t_r$  applied before, so that the equation obtained by summing (57) over all  $r$  goes over into (52) on putting all the  $t_r$  equal to  $t$ . To verify the second, we note that we may take  $\Psi$  with each  $t_r$  put equal to some given  $t$  to be arbitrary, and equations (57) will then have a solution provided they are consistent, in the sense that  $\partial/\partial t_s$  of  $\partial\Psi/\partial t_r$  given by one of equations (57) equals  $\partial/\partial t_r$  of  $\partial\Psi/\partial t_s$  given by another, for every pair  $r, s$ . The condition for this consistency is easily seen to be that all the operators  $H_r^*(t_r)$  shall commute with each other. We see from (24) and (26) that they do commute provided

$$(t_r - t_s)^2 < (\mathbf{x}_r - \mathbf{x}_s)^2 \quad (58)$$

for every pair  $r, s$ . Thus we have the conditions (58) putting a restriction on the domain of existence of the wave function  $\Psi$ , and inside this domain of existence we can obtain a  $\Psi$  corresponding to any  $\psi^*$  that satisfies (52).

The restriction (58) on the domain of existence of  $\Psi$  is to be expected also from the physical interpretation of the wave function. The natural interpretation to assume for the wave function  $\Psi$ , as a generalization of that for the wave function  $\psi$  or  $\psi^*$ , is that the square of its modulus for any set of values for the  $\mathbf{x}_r, t_r$  is proportional to the probability of each of the particles being in a small volume about the point  $\mathbf{x}_r$  at the time  $t_r$ , with the field in a specified state (i.e.



with specified photons in existence). Such an interpretation would not be permissible outside the region (58), because of the interference that there would then be between the observations of the positions of the various particles at the various times.

The scheme of wave equations (57) is assumed to describe completely the interaction between the various charged particles and the field and should therefore include the interaction between one charged particle and another, since in field theory there is no direct interaction between one particle and another but only an indirect one, arising from each particle influencing the field in its neighbourhood and this influence spreading out till it reaches the other particles. Thus these equations should include forces between the charged particles of a type which reduces to the Coulomb forces in non-relativistic approximation. It is not at all evident that the equations do include such forces, since they appear to take into account the action of the field on the particles, but not the action of the particles on the field. In order to verify that they are complete, we must go back to the Heisenberg picture and see that the equations of motion of the field are then analogous to the classical equations of motion of the field with the influence of the charged particles duly taken into account.

Going back to the Heisenberg picture requires us to put all the time variables equal and to take the Hamiltonian (48). With this Hamiltonian we get as the equation of motion for  $\partial A_\mu / \partial t$

$$\begin{aligned} \frac{\partial^2 A_\mu}{\partial t^2} &= \left[ \frac{\partial A_\mu}{\partial t}, H_F + \sum_r H_r \right] \\ &= \nabla^2 A_\mu + \sum_r \left[ \frac{\partial A_\mu}{\partial t}, H_r \right] \end{aligned} \quad (59)$$

from (47). But from (49) and (30)

$$\left[ \frac{\partial A_0}{\partial t}, H_r \right] = e_r \left[ \frac{\partial A_0}{\partial t}, A_{0r} \right] = 4\pi e_r \delta(\mathbf{x} - \mathbf{x}_r)$$

and similarly

$$\left[ \frac{\partial \mathbf{A}}{\partial t}, H_r \right] = e_r \left[ \frac{\partial \mathbf{A}}{\partial t}, (\boldsymbol{\alpha}_r, \mathbf{A}_r) \right] = -4\pi e_r \boldsymbol{\alpha}_r \delta(\mathbf{x} - \mathbf{x}_r).$$

Thus (59) becomes

$$\left. \begin{aligned} \frac{\partial^2 A_0}{\partial t^2} - \nabla^2 A_0 &= 4\pi \sum_r e_r \delta(\mathbf{x} - \mathbf{x}_r) \\ \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} &= -4\pi \sum_r e_r \boldsymbol{\alpha}_r \delta(\mathbf{x} - \mathbf{x}_r). \end{aligned} \right\} \quad (60)$$



These are the equations required by Maxwell's theory for charges  $e_r$  at the various points  $\mathbf{x}_r$  moving with the velocities  $-\alpha_r$ , which are the velocities required by the relativistic theory of the electron. Hence the action of the particles on the field, giving rise to Coulomb forces between the particles in non-relativistic approximation, is correctly taken into account in the Hamiltonian (48), and thus also in the scheme of equations (57).

To complete the theory, we must now obtain the supplementary conditions to go with the wave equations (57). The conditions which naturally suggest themselves are (39) with  $\Psi$  instead of  $\psi$ . These would be equivalent to (36) with the constant operators  $A_\mu$  of the Schrödinger picture replaced by  $A_\mu^*$ 's and  $\Psi$  instead of  $\psi$ . These conditions need amendment, though, as may be seen from the following considerations. Equations (57) may be regarded as supplementary conditions and have to be consistent with any further supplementary conditions, in the way discussed in connexion with equations (37). (This consistency requirement is equivalent to the requirement that, if the further supplementary conditions hold for some value for each of the  $t_r$ , they shall hold generally.) Now the operators in (57) are built up from

$$\mathbf{p}_r - e_r \mathbf{A}_r^*(t_r), \quad W_r - e_r A_{0r}^*(t_r), \quad (61)$$

$W_r$  meaning the energy operator of the  $r$ th particle, i.e.  $i\hbar\partial/\partial t_r$  when operating to the right. We have, remembering that our  $A^*$ 's satisfy the same commutability relations as the  $A$ 's in the Heisenberg picture with no charged particles present, namely equations (24), (27), (28) and (29),

$$\begin{aligned} \left[ \frac{\partial A_0^*}{\partial t} + \text{div } \mathbf{A}^*, W_r - e_r A_{0r}^*(t_r) \right] &= -e_r \left[ \frac{\partial A_0^*}{\partial t}, A_{0r}^*(t_r) \right] \\ &= -e_r \frac{\partial}{\partial t} [A_0^*, A_{0r}^*(t_r)] = \mp 2e_r \frac{\partial}{\partial t} \delta\{(\mathbf{x} - \mathbf{x}_r)^2 - (t - t_r)^2\}, \end{aligned}$$

the minus or plus sign being taken according to whether  $t > t_r$  or  $t < t_r$ . Thus

$$\begin{aligned} \left[ \frac{\partial A_0^*}{\partial t} + \text{div } \mathbf{A}^*, W_r - e_r A_{0r}^*(t_r) \right] &= \pm 2e_r \frac{\partial}{\partial t_r} \delta\{(\mathbf{x} - \mathbf{x}_r)^2 - (t - t_r)^2\} \\ &= \mp 2e_r [\delta\{(\mathbf{x} - \mathbf{x}_r)^2 - (t - t_r)^2\}, W_r - e_r A_{0r}^*(t_r)], \end{aligned}$$

so that

$$\frac{\partial A_0^*}{\partial t} + \text{div } \mathbf{A}^* + 2 \sum_s \pm e_s \delta\{(\mathbf{x} - \mathbf{x}_s)^2 - (t - t_s)^2\} \quad (62)$$

commutes with  $W_r - e_r A_{0r}^*(t_r)$ , the plus or minus sign being taken for



each term in the sum according to whether  $t > t_s$  or  $t < t_s$ . Similarly (62) commutes with all the other quantities (61). We therefore take as our supplementary conditions

$$\left\{ \frac{\partial A_0^*}{\partial t} + \operatorname{div} \mathbf{A}^* + 2 \sum_s \pm e_s \delta\{(\mathbf{x} - \mathbf{x}_s)^2 - (t - t_s)^2\} \right\} \Psi = 0. \quad (63)$$

There is one of these conditions for each point  $\mathbf{x}$ ,  $t$  in space-time, the  $\mathbf{x}$ ,  $t$  variables being quite arbitrary and independent of all the  $\mathbf{x}_s$ ,  $t_s$  variables. These supplementary conditions are consistent with the wave equations (57), since the operator (62) commutes (for all values of  $\mathbf{x}$ ,  $t$ ) with the operators in (57). The additional terms in (63), involving the  $\delta$  function, are necessary to secure this consistency. These additional terms do not interfere with the mutual consistency of the various equations (63) obtained by giving different values to  $\mathbf{x}$ ,  $t$ , since they commute with all the other operators in these equations.

## 80. The Quantization of Electron Waves

If the charged particles in the preceding section are electrons, we should have to impose the additional condition on  $\Psi$  that it shall be antisymmetrical between all the electrons. We can then put the equations for  $\Psi$  into a more concise and more symbolic form, by the use of a procedure of quantization of electron wave functions, which was discovered by Jordan and Wigner.<sup>†</sup> This procedure is the analogue for particles satisfying the exclusion principle of the second quantization discussed in § 62 for particles satisfying the Einstein-Bose statistics, and we shall deal with it on corresponding lines to those used in § 62 for the Einstein-Bose case.

We begin by describing the states of our assembly of particles by antisymmetrical representatives  $(q_1 q_2 \dots q_n |)$ , of the kind we had in § 57. We introduce the observables  $n_a$  having the same meaning as in § 62, i.e.  $n_1$  is the number of  $q$ 's equal to  $q^{(1)}$ ,  $n_2$  the number equal to  $q^{(2)}$ , and so on. Each of these  $n$ 's now has as eigenvalues only 0 and 1, since for any  $n$  having a value greater than 1 the antisymmetrical wave function  $(q_1 q_2 \dots q_n |)$  would vanish identically. We pass over to the representatives  $(n_1 n_2 \dots |)$ , assuming as the connexion between the two representatives of any state

$$(n_1 n_2 \dots |) = \pm (q_1 q_2 \dots q_n |). \quad (64)$$

This equation is the analogue of (2) of Chapter XI. The normalizing

<sup>†</sup> Jordan and Wigner, *Z. f. Physik*, **47** (1928), 631.



factor  $[n!/n_1!n_2!n_3!\dots]^{\frac{1}{2}}$  in (2) of Chapter XI is not required in (64) on account of the eigenvalues of the  $n$ 's now being restricted to 0 and 1. We need, however, a  $\pm$  sign in (64), which we did not have in (2) of Chapter XI, since for given values of the  $n$ 's, the values of the  $q$ 's in  $(q_1q_2\dots q_n|)$  are fixed but not their order, so that,  $(q_1q_2\dots q_n|)$  being antisymmetrical, there will be an ambiguity in its sign. We must set up a rule for specifying the sign in any particular case. We can do this by arranging all the eigenvalues of a  $q$  arbitrarily in some definite order, say the order

$$q^{(1)}, q^{(2)}, q^{(3)}, \dots, \quad (65)$$

which may conveniently be taken the same as the order in which the  $n$ 's are written in  $(n_1n_2\dots|)$ , and then requiring that the  $+$  sign shall be taken when the  $q$ 's in (64), which form a selection from the total set (65), can be brought into the order in which they appear in (65) by an even number of interchanges, and the  $-$  sign otherwise.

We must now obtain the transformation law for the representative of a dynamical variable  $U$ , of the form of (3) of Chapter XI, from the  $q$ -representation to the  $n$ -representation. Following through the same method as in Chapter XI and transforming the equation

$$\psi_2 = U\psi_1, \quad (66)$$

we obtain, corresponding to equation (8) of Chapter XI,

$$(n_1n_2\dots|2) = \sum_a n_a U_{aa}(n_1n_2\dots|1) + \sum_a \sum_{b \neq a} \pm U_{ab}(n_1n_2\dots n_a-1\dots n_b+1\dots|1), \quad (67)$$

where  $(n_1n_2\dots n_a-1\dots n_b+1\dots|1)$  is understood to be zero if either  $n_a-1$  or  $n_b+1$  is not 0 or 1. With regard to the ambiguity of sign occurring in (67), we must take the  $-$  sign in those cases where there is a  $-$  sign in one and only one of the equations

$$(n_1n_2\dots|1) = \pm(q_1q_2\dots q_n|1) \quad (68)$$

$$(n_1n_2\dots n_a-1\dots n_b+1\dots|1) = \pm(q_1q_2\dots q^{(b)} \text{ for } q^{(a)}\dots q_n|1).$$

It is easily seen that this condition for the  $-$  sign is the same as the condition that the number of  $q$ 's mentioned on the right-hand side of (68) that lie between  $q^{(a)}$  and  $q^{(b)}$  in the sequence (65) shall be odd, or that  $\sum_c n_c$ , where the summation is taken over all  $c$  for which  $q^{(c)}$  lies between  $q^{(a)}$  and  $q^{(b)}$  in the sequence (65), shall be odd. When  $\sum_c n_c$  is even, we must have the  $+$  sign.

If we take any  $n_a$  and form  $1-2n_a$ , we get an observable having as



eigenvalues 1 and  $-1$ , and thus of the same nature as the  $\sigma$ 's dealt with in § 19. Let us put

$$1 - 2n_a = \sigma_{za} \quad (69)$$

and introduce the  $\sigma_{xa}$  and  $\sigma_{ya}$  that are associated with it. Then in a representation with  $\sigma_{za}$  diagonal,  $\frac{1}{2}(\sigma_{xa} - i\sigma_{ya})$  and  $\frac{1}{2}(\sigma_{xa} + i\sigma_{ya})$  will be represented, according to (57) of § 19, by

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad (70)$$

respectively, and will thus be to a certain extent analogous to the  $e^{iwa}$  and  $e^{-iwa}$  or to the  $\bar{\xi}_a$  and  $\xi_a$  of § 62. There will be one set of  $\sigma_{xa}$ ,  $\sigma_{ya}$  and  $\sigma_{za}$  for each  $a$ , and members of one set will commute with members of any other set.

The form of the representatives (70) shows that when  $\frac{1}{2}(\sigma_{xa} - i\sigma_{ya})$  or  $\frac{1}{2}(\sigma_{xa} + i\sigma_{ya})$  is multiplied into a  $\psi$  whose representative is  $(n_1 n_2 \dots n_a \dots |)$ , the representative of the product is  $(n_1 n_2 \dots n_a - 1 \dots |)$  or  $(n_1 n_2 \dots n_a + 1 \dots |)$  respectively. Hence equation (67) is the representative of

$$\psi_2 = \sum_a n_a U_{aa} \psi_1 + \sum_a \sum_{b \neq a} \pm U_{ab} \frac{1}{2}(\sigma_{xa} - i\sigma_{ya}) \frac{1}{2}(\sigma_{xb} + i\sigma_{yb}) \psi_1.$$

Since this holds whenever (66) holds, we must have

$$U = \sum_a n_a U_{aa} + \sum_a \sum_{b \neq a} \pm \frac{1}{2}(\sigma_{xa} - i\sigma_{ya}) U_{ab} \frac{1}{2}(\sigma_{xb} + i\sigma_{yb}). \quad (71)$$

In order to get rid of the  $\pm$  sign in this result, we introduce the dynamical variables

$$\xi_b = \sigma_{z1} \sigma_{z2} \sigma_{z3} \dots \sigma_{z, b-1} \frac{1}{2}(\sigma_{xb} + i\sigma_{yb}), \quad (72)$$

where the product of  $\sigma_z$ 's consists of the  $\sigma_z$ 's corresponding to all the  $q$ 's in the sequence (65) up to  $q^{(b-1)}$ . The conjugate complex of  $\xi_a$  is

$$\bar{\xi}_a = \frac{1}{2}(\sigma_{xa} - i\sigma_{ya}) \sigma_{z1} \sigma_{z2} \sigma_{z3} \dots \sigma_{z, a-1}. \quad (73)$$

We now have for  $b \neq a$ , since the square of any  $\sigma_z$  is unity,

$$\bar{\xi}_a \xi_b = \frac{1}{2}(\sigma_{xa} - i\sigma_{ya}) \left\{ \begin{matrix} \sigma_{za} \sigma_{z, a+1} \sigma_{z, a+2} \dots \sigma_{z, b-1} \\ \sigma_{z, a-1} \sigma_{z, a-2} \dots \sigma_{z, b+1} \sigma_{zb} \end{matrix} \right\} \frac{1}{2}(\sigma_{xb} + i\sigma_{yb}), \quad (74)$$

where the first or second line in the  $\{ \}$  brackets is to be taken, according to whether  $q^{(a)}$  comes before or after  $q^{(b)}$  in the sequence (65). From (55) of Chapter III,

$$(\sigma_{xa} - i\sigma_{ya}) \sigma_{za} = \sigma_{xa} - i\sigma_{ya}$$

and

$$\sigma_{zb} (\sigma_{xb} + i\sigma_{yb}) = \sigma_{xb} + i\sigma_{yb}.$$



Thus the  $\sigma_{za}$  and  $\sigma_{zb}$  factors in the  $\{ \}$  brackets in (74) may be omitted, leaving in these brackets

$$\left\{ \begin{array}{l} \sigma_{z, a+1} \sigma_{z, a+2} \dots \sigma_{z, b-1} \\ \sigma_{z, a-1} \sigma_{z, a-2} \dots \sigma_{z, b+1} \end{array} \right\},$$

or, from (69),

$$\left\{ \begin{array}{l} (1-2n_{a+1})(1-2n_{a+2}) \dots (1-2n_{b-1}) \\ (1-2n_{a-1})(1-2n_{a-2}) \dots (1-2n_{b+1}) \end{array} \right\}. \quad (75)$$

The operator in the  $\{ \}$  brackets now commutes with  $\frac{1}{2}(\sigma_{xb} + i\sigma_{yb})$  and may be taken to the right of  $\frac{1}{2}(\sigma_{xb} + i\sigma_{yb})$  in (74), and when multiplied into a  $\psi$  represented by  $(n_1 n_2 \dots |)$ , will be equivalent to the factor  $\pm 1$ , the  $+$  or  $-$  sign being taken according to whether  $\sum_c n_c$ , summed for all  $c$  for which  $q^{(c)}$  lies between  $q^{(a)}$  and  $q^{(b)}$  in the sequence (65), is even or odd. This holds in both cases—when  $q^{(a)}$  comes before  $q^{(b)}$  and we have to take the first line in (75), and when  $q^{(a)}$  comes after  $q^{(b)}$  and we have to take the second. Thus the operator (75) is equivalent to the  $\pm$  sign in (67) and (71), and (71) reduces to

$$U = \sum_a n_a U_{aa} + \sum_a \sum_{b \neq a} \bar{\xi}_a U_{ab} \xi_b, \quad (76)$$

from which the ambiguity of sign has disappeared.

We must now determine the commutability relations for the  $\xi$ 's and  $\bar{\xi}$ 's. We shall first prove that

$$\xi_a \xi_b + \xi_b \xi_a = 0. \quad (77)$$

If  $a$  and  $b$  are different, suppose  $q^{(a)}$  comes before  $q^{(b)}$  in the sequence (65). Then one of the factors in the expression for  $\xi_b$  given by formula (72) anticommutes with one of the factors in the expression for  $\xi_a$ , namely  $\sigma_{za}$  in  $\xi_b$  anticommutes with  $(\sigma_{xa} + i\sigma_{ya})$  in  $\xi_a$ , but apart from this every factor in  $\xi_b$  commutes with every factor in  $\xi_a$ . Thus  $\xi_b$  must anticommute with  $\xi_a$ . If  $a$  and  $b$  are the same, equation (77) states that  $\xi_a^2 = 0$ , and this holds since  $(\sigma_{xa} + i\sigma_{ya})^2 = 0$ . Thus equation (77) holds generally. In a similar way we can show that

$$\bar{\xi}_a \bar{\xi}_b + \bar{\xi}_b \bar{\xi}_a = 0 \quad (78)$$

$$\text{and} \quad \xi_a \bar{\xi}_b + \bar{\xi}_b \xi_a = 0 \quad \text{for } b \neq a. \quad (79)$$

We have further

$$\begin{aligned} \bar{\xi}_a \xi_a &= \frac{1}{2}(\sigma_{xa} - i\sigma_{ya}) \frac{1}{2}(\sigma_{xa} + i\sigma_{ya}) \\ &= \frac{1}{2}(1 - \sigma_{za}) \\ &= n_a, \end{aligned} \quad (80)$$



from (69), and again

$$\begin{aligned}\xi_a \bar{\xi}_a &= \frac{1}{2}(\sigma_{xa} + i\sigma_{ya}) \frac{1}{2}(\sigma_{xa} - i\sigma_{ya}) \\ &= \frac{1}{2}(1 + \sigma_{za}) \\ &= 1 - n_a.\end{aligned}\tag{81}$$

Equations (80) and (81) show that

$$\xi_a \bar{\xi}_a + \bar{\xi}_a \xi_a = 1$$

and thus that (79) can be extended to

$$\xi_a \bar{\xi}_b + \bar{\xi}_b \xi_a = \delta_{ab}.\tag{82}$$

The quantum conditions (77), (78) and (82) are to be compared with (15) of Chapter XI for the Einstein-Bose case. The only difference is a change in the signs on the left-hand sides.

From (80), (76) can be expressed in the form

$$U = \sum_{ab} \bar{\xi}_a U_{ab} \xi_b,\tag{83}$$

which is the same as (16) of Chapter XI. If we suppose  $U$  to be the Hamiltonian of the assembly of particles, we shall get as the equations of motion for the  $\xi$ 's,

$$\begin{aligned}i\hbar \dot{\xi}_a &= \xi_a U - U \xi_a \\ &= \sum_{cb} (\xi_a \bar{\xi}_c U_{cb} \xi_b - \bar{\xi}_c U_{cb} \xi_b \xi_a) \\ &= \sum_{cb} (\xi_a \bar{\xi}_c + \bar{\xi}_c \xi_a) U_{cb} \xi_b\end{aligned}$$

from (77). This reduces, with the help of (82), to

$$i\hbar \dot{\xi}_a = \sum_b U_{ab} \xi_b,\tag{84}$$

which is of the same form as (21) of Chapter XI, and as the wave equation for a single one of our particles by itself, with  $\xi_a$  playing the part of  $(q^{(a)}|)$ . Thus our present scheme of equations may, like the scheme of § 62, be considered as coming from a process of second quantization, the quantum conditions for which are (77), (78), and (82).

We may apply the foregoing scheme to the problem of several electrons interacting with the electromagnetic field. We may take the operator on the right-hand side of (52) as the above Hamiltonian  $U$ . The equation of motion (84) will then be of the same form as one of the equations (57), with  $\Psi$  involving only one set of space-time variables  $\mathbf{x}, t$  and with a second quantization applied to  $\Psi$ , so that its values for different values of  $\mathbf{x}$  are not numbers, but operators



satisfying the quantum conditions (77), (78), and (82) (the last of which has to be rewritten with a  $\delta$  function instead of the two-suffix  $\delta$  symbol since the eigenvalues of  $\mathbf{x}$  take on continuous ranges of values).

## 81. Conclusion

The foregoing theory provides a quantum electrodynamics which is a satisfactory analogue of classical electrodynamics, each of the features of classical electrodynamics having its quantum counterpart. As a description of nature, though, the theory is incomplete, as it suffers from the same limitations as the classical theory with regard to the distribution of electric charge inside an electron. The quantum equations we have discussed for electrons in interaction with the field correspond to classical equations based on the point-charge model of the electron, i.e. the model in which all the charge is assumed to be concentrated at one point. Such a model in classical electrodynamics leads to an infinite mass for the electron, since the energy-density in the neighbourhood of the point where the charge is situated tends to infinity in a way that makes the total energy non-convergent. Analogously, the quantum theory that we have set up also leads to an infinite mass for the electron. This infinite mass here shows itself through the expression for  $d\psi/dt$  given by the wave equation (50) having an infinite value, owing to the non-convergence of the contributions to  $d\psi/dt$  arising from terms in the Hamiltonian corresponding to Fourier components of the field with very short wave-lengths. In consequence, the wave equation (50) and its equivalents (52) and (57) do not strictly have any solutions.

The theory can be made to give finite and sensible answers for elementary problems, such as the emission and absorption of radiation whose wave-length is not too short, since it allows the probabilities one wishes to calculate to be expressed in terms of semi-convergent series or integrals, in which one can simply ignore the divergent part arising from the short wave-lengths. Such a procedure can lead to a definite answer, of course, only when the divergent part is clearly separated from the part we are interested in. The condition for this is that the important wave-lengths for the problem under consideration shall be long compared with the classical radius of the electron. The limitations in the applicability of quantum electrodynamics thus correspond precisely to those of classical electro-



dynamics. The amendments required in classical theory in order to make it apply accurately to the elementary charged particles are thus not provided by the passage to the quantum theory, that is, by the taking into considerations of the disturbances accompanying measurements. It seems that some essentially new physical ideas are here needed.







## INDEX OF DEFINITIONS

- action variable, 135.
- angle variable, 136.
- angular momentum, 145.
- anticommute, 67.
- antisymmetrical wave function, 211.
  
- bar notation, 21.
- basic  $\psi$ 's, 49.
- belonging to an eigenvalue, 32.
- Bohr's frequency condition, 122, 179.
- boundary condition, 155.
- bracket notation, 49.
  
- character (of a group), 218.
- class of permutations, 215.
- closed state, 155.
- commutability relation, 88.
- commute, 27.
- compatible observations, 48.
- complete set of commuting observables, 58.
- conjugate complex, 22.
- — of a linear operator, 53.
- imaginary, 22.
- constant of the motion, 121.
- contact transformation, 111.
  
- de Broglie waves, 132.
- degenerate system, 172.
- dependent states, 15.
- diagonal element, 25, 80.
- matrix, 56, 80.
- $\delta_{rs}$ , 50.
- $\delta$  function, 72.
  
- eigen, 32.
- eigenfunction, 100.
- Einstein-Bose statistics, 212.
- electric density, 256.
- element of a matrix, 25.
- exclusion principle, 212.
- exclusive sets of states, 217.
  
- Gibbs ensemble, 139.
  
- $h$ ,  $\hbar$ , 90.
- half-width of absorption line, 206.
  
- Hamiltonian, 116.
- Heisenberg picture, 118.
- Heisenberg representation, 121.
- Hermitian, 29.
  
- identical permutation, 213.
- improper function, 72.
- independent states, 15.
  
- Kramers-Heisenberg dispersion formula, 247.
  
- Landé's formula, 186.
- length of a vector, 23.
  
- matrix, 25, 79.
- mixed representative, 64.
- multiplet, 184.
  
- non-degenerate system, 172.
- normalization, 24.
- with continuous parameter, 78.
  
- observable, 24, 37.
- having a value, 43.
- having an average value, 44.
- orthogonal, 24.
- orthogonality theorem, 33.
  
- Pauli's exclusion principle, 212.
- permutations, 220.
- phase factor, 24.
- space, 139.
- Planck's constant, 90.
- Poisson bracket, 89.
- positive square root, 42.
- positron, 271.
- probability amplitude, 66.
- of observable having a value, 44.
- proper-energy, 181.
  
- quantum condition, 88.
  
- reciprocal of an observable, 40.
- permutation, 214.
- representation, 52.
- representative, 52.



- scatterer, 187.  
Schrödinger picture, 118.  
Schrödinger's wave equation, 116.  
second quantization, 234.  
selection rule, 158.  
similar permutations, 214.  
Sommerfeld's formula, 269.  
spacial quantization, 150.  
square root of an observable, 41.  
state, 11, 18.  
— of absorption, 189.  
— of motion, 7.  
— of polarization, 5.  
stationary state, 117.  
stimulated emission, 179.  
supplementary condition, 283.  
symmetrical wave function, 210.  
uncertainty principle, 104.  
unit matrix, 51, 80.  
unitary, 111.  
wave equation, 116.  
— function, 116.  
— packet, 103.  
weight function, 85.  
well-ordered function, 113.  
zero state, 237.